# The Balassa-Samuelson Effect in Slovenia: an Olley-Pakes Total Factor Productivity Approach

Sjoerd Anton van der Schaar

Supervised by Prof. Dr. Lorenzo Pozzi

Erasmus School of Economics, Erasmus University Rotterdam

# The Balassa-Samuelson Effect in Slovenia: an Olley-Pakes Total Factor Productivity Approach

**Abstract**

This paper introduces Olley-Pakes productivity estimates into the well-known Balassa-Samuelson framework. The Balassa-Samuelson theorem as developed by Balassa (1964) and Samuelson (1964) states that productivity differentials between the tradable and non-tradable sector are the main drivers for an appreciating real exchange rate. Previous empirical studies on the Balassa-Samuelson effect implemented productivity measures which suffer from econometric issues, most notably simultaneity. I correct for simultaneity by using the Olley-Pakes method of productivity estimation. Firm level data provided by the Slovenian Statistical Office allowed me to reliably estimate productivity differentials. I find evidence of a present bias where productivity differentials are typically overestimated if simultaneity has not been accounted for. Following this result I was able to find evidence of a present Balassa-Samuelson effect in levels. However, due to non-stationarity I implemented first differences where I was not able to find cointegration. Consequently, I conclude that the Balassa-Samuelson effect is not as prevalent when econometrically sound productivity measures are implemented.

## Table of contents

# 1. Introduction

Over the last twenty years the availability of firm level data has increased significantly. This has opened many doors to academic researchers to study firm behavior on a micro level. Hand in hand with the above came new methodology to reliably estimate a firm's productivity. Most notably Olley and Pakes in 1994 with their paper on the U.S. telecommunications industry.

One of the doors firm-level data can open relates to the Balassa-Samuelson theorem. Balassa (1964) and Samuelson (1964) laid out the well-known framework where productivity differentials between the tradable and non-tradable sectors explain systematic variation in relative prices and therefore real exchange rates. It is generally observed that productivity growth in the tradable goods producing sector exceeds that of the non-tradable goods producing sector, this especially holds truth in transition economies. The prices of tradable goods are equalized by the law of one price, which does not hold for the non-tradable goods producing sector. Given the higher productivity in the tradable goods producing sector, wages will rise in both sectors if we assume labor to be domestically mobile across sectors. Firms in the non-tradable sector will only be able to pay these higher wages by increasing the relative price of non-tradable goods. Countries with higher productivity growth levels in the tradable goods producing sector will, ceteris paribus, have higher relative wages and therefore higher relative prices of non-tradable goods. Ultimately resulting in an appreciation of the real exchange rate.

The Balassa-Samuelson theorem has been thoroughly empirically analyzed in the 1990's and early 2000's. Early empirical literature on the Balassa Samuelson hypothesis often found convincing estimates (Rother, 2000; Halpern and Wyplosz, 2001; De Broeck and Sløk, 2001). The impact of productivity differentials on domestic inflation and the real exchange rate was found to be around one-to-one. More recent empirical studies questioned the magnitude of the hypothesis (Égert, 2003; Égert et al, 2006; Mihaljek and Klau, 2008) with estimates ranging from 0 to 0.5. They were especially skeptical of the theorem's assumptions. Assumptions that are unlikely to hold in the real world. For instance, Mihaljek and Klau (2008) found little significant evidence for a Balassa-Samuelson effect in the short run despite finding it in the long run. Possibly indicating that the assumption of perfect capital and labor mobility does not hold in the short run.

Most empirical studies on the Balassa-Samuelson hypothesis glance over the issues related to the productivity measures used in their estimations. Productivity differences are estimated by means of labor productivity or simple Solow residuals. Both methods of productivity estimation suffer from econometric issues. Consequently, these productivity measures produce biased estimates of the Balassa-Samuelson effect. Labor productivity as an approximation of productivity ignores other factors of production which enhance production and therefore productivity as well. In this setting, productivity differences are typically overestimated when labor productivity is used. On the other hand, Solow residuals provide a more suitable option. Solow residuals follow a production function and consider multiple factors of production. Issues arise when endogeneity of productivity is considered. Factors of production are partly determined by the firm's previous knowledge of its productivity. Knowledge that I, as a researcher, do not have access to which biases my estimates.

This paper attempts to tackle these estimations issues by introducing the Olley-Pakes method of productivity estimation into the Balassa-Samuelson framework. This method of estimation incorporates a firm's investment decision in order to proxy for unobserved productivity shocks. Olley-Pakes productivity estimation requires a firm-level dataset to capture firm decisions. The Slovenian statistical office has thankfully granted me access to such a database. The database comprises financial records of Slovenian firms over a period of 13 years. This is the only paper to my knowledge that has attempted to incorporate Olley-Pakes estimates of productivity into the Balassa-Samuelson framework. It is therefore the main contribution of this paper to the existing literature covering the theorem.

In this paper I find evidence of a present bias if one used simple Solow residuals as a measure for productivity. Olley-Pakes productivity indices indicated smaller productivity gaps between the tradable and non-tradable sector. Hence, Solow residuals would overestimate the Balassa-Samuelson effect. Using the Olley-Pakes productivity indices in the Balassa-Samuelson framework resulted in the following estimates (Table 1) in levels, first differences and additional controls. Similar to previous studies, I was able to find a sizable Balassa-Samuelson effect in levels. However, stationarity tests revealed non-stationarity in levels for both the dependent and independent variable. Stationarity was found in first differences where the Balassa-Samuelson effect remained significant and positive albeit losing its magnitude. I was not able to find cointegration in the first differences specifications. The results I found resemble the more recent literature where researchers are rather skeptic of a sizable presence of the Balassa-Samuelson effect.

Table 1: Estimates of the Balassa-Samuelson effect for Slovenia in levels, first differences and additional controls. 1995-2006.

| Specification | $\beta$ |
|---|---|
| $q = \lambda + \beta\big((\overline{\omega}_t^T - \overline{\omega}_t^{NT}) - (\overline{\omega}_t^{T*} - \overline{\omega}_t^{NT*})\big) + \mu_t$ | 0.5561<br>(0.1390) |
| $\Delta q = \lambda + \beta * \Delta\big((\overline{\omega}_t^T - \overline{\omega}_t^{NT}) - (\overline{\omega}_t^{T*} - \overline{\omega}_t^{NT*})\big) + \mu_t$ | 0.0324<br>(0.0072) |
| $\Delta q = \lambda + \beta * \Delta\big((\overline{\omega}_t^T - \overline{\omega}_t^{NT}) - (\overline{\omega}_t^{T*} - \overline{\omega}_t^{NT*})\big) + \gamma * \Delta\phi + \mu_t$ | 0.0290<br>(0.0064) |

*All coefficients reported are highly significant, standard errors are shown in brackets. The specifications are taken from eq.(23), eq.(24) and eq.(25).*

*Source: Annual reports of companies and sole proprietors, SURS, 2018; OECD, 2018; Banka Slovenije, 2018; Deutsche Bundesbank, 2018*

In terms of structure, there is a division of all sections into two parts. On the one hand covering the Balassa Samuelson theorem and the other focusing on productivity. The following is structured in a traditional manner where section 2 will review previous literature and the theoretical framework related to the Balassa-Samuelson theorem and productivity. Section 3 and 4 cover methodology and the data used to implement said methodology. Followed by sections 5 and 6 which will comprise of results and discussion of said results. The paper will be concluded by section 7.

## 2. Literature review

### 2.1 Balassa Samuelson hypothesis

The Balassa-Samuelson hypothesis in short states that disproportionate productivity gains in the tradable sector versus the non-tradable sector results in an increase in inflation and thus the real exchange rate. This result is analytically shown in the following section. Empirics covering the Balassa-Samuelson hypothesis is discussed in the section after the theoretical model.

### 2.1.1 Theoretical model

In order to illustrate the Balassa Samuelson hypothesis mathematically I first consider the domestic Cobb-Douglas functions for the two sectors:

$$Y_T = A_T L_T^{\gamma} K_T^{1-\gamma} \quad (1)$$

$$Y_{NT} = A_{NT} L_{NT}^{\delta} K_{NT}^{1-\delta} \quad (2)$$

where $Y$ represents output and $A$, $L$ and $K$ constitute productivity, labor input and capital input respectively. The subscripts $T, NT$ stand for the tradable and non-tradable goods producing sectors, $\gamma$ and $\delta$ represent the intensity of labor input in the production process. Constant returns to scale are considered therefore $0 < \gamma < 1$ and $0 < \delta < 1$. Labor intensity is often observed to be higher in the non-tradable sector (mainly services, $\delta > \gamma$). However, for simplicity reasons I assume $\delta = \gamma$ for now.

Capital is assumed to be perfectly mobile across sectors and also internationally. Hence the interest rate is determined in the world market and thus exogenous to the model. Labor is domestically mobile across sectors, however labor is not mobile internationally. Therefore the wage rate is determined domestically and thus endogenous to the model. Given perfect competition in both sectors we obtain the following FOC's for profit maximization:

$$\frac{i}{P_T} = (1-\gamma) A_T \left(\frac{K_T}{L_T}\right)^{-\gamma} \quad (3) \qquad \frac{w}{P_T} = \gamma A_T \left(\frac{K_T}{L_T}\right)^{1-\gamma} \quad (4)$$

$$\frac{i}{P_{NT}} = (1-\delta) A_{NT} \left(\frac{K_{NT}}{L_{NT}}\right)^{-\delta} \quad (5) \qquad \frac{w}{P_{NT}} = \delta A_{NT} \left(\frac{K_{NT}}{L_{NT}}\right)^{1-\delta} \quad (6)$$

where $i$ and $w$ are the exogenously determined interest rate and the endogenously determined wage rate. $P_T$ is the price level for tradable goods whereas $P_{NT}$ is the price level for non-tradable goods. Equations (3) to (6) represent marginal product equals marginal cost conditions.

The first key insight can be seen in equations (4) and (6). If productivity growth is higher in the tradable sector ( $\Delta A_T > \Delta A_{NT}$) and wages develop equally across sectors, then the relative price of non-tradable goods $\left(\frac{P_{NT}}{P_T}\right)$ has to increase in order to equalize[1] (holding all else constant). This is also known as the internal transmission mechanism where productivity gains in the tradable sector lead to an increase in prices of non-tradable goods. The producers in the tradable goods sector can only utilize the wage channel to compensate for its productivity gains as it cannot alter the global interest rate ($i$) and/or the price of tradable goods ($P_T$).

Given the assumption where labor intensity is assumed to be equal ($\delta = \gamma$) across the two sectors, it is reasonable to assume that the capital over labor ratio is equal across sectors. Log-differentiating equations (4) and (6) will then yield a more compact and clear expression for the internal transmission mechanism:

(7) $$p_{NT} - p_T = \frac{\delta}{\gamma}(a_T - a_{NT})$$

Note that the relative price of non-tradable goods can also increase if labor intensity is assumed to be higher in the non-tradable sector ($\delta > \gamma$) and productivity growth is similar. If we assume labor intensity to be equal ($\delta = \gamma$) then the relationship between the relative price of non-tradables and productivity growth will be one-to-one.

Extending the model to a two country setting where the same assumptions hold and parameters share identical values, I obtain the following aggregate price levels for both domestic and abroad (abroad denoted by *).

(8) $$p = \alpha p_T + (1 - \alpha) p_{NT}$$

(9) $$p^* = a^* p_T^* + (1 - \alpha^*) p_{NT}^*$$

where $p$ and $p^*$ denote domestic and foreign aggregate price levels. $\alpha$ and $a^*$ represent the share of consumption on traded goods.

The real exchange rate ($q$) is denoted by the following equation:

(10) $$q = (e + p^*) - p$$

---

[1] $P_T \gamma A_T (\frac{K_T}{L_T})^{1-\gamma} = w = P_{NT} \delta A_{NT} (\frac{K_{NT}}{L_{NT}})^{1-\delta}$ , $\Delta A_T > \Delta A_{NT}$ therefore $\Delta P_{NT} > \Delta P_T$ is necessary to equalize the condition where wages equalize across sectors.

where $e$ is the nominal exchange rate defined as the amount of domestic currency per unit of foreign currency. Substituting (8),(9) into (10) and adding $p_T - p_T^*$ to both sides of (10) the following is obtained:

$$q = (e + p_T^* - p_T) + (1 - \alpha^*)(p_{NT}^* - p_T^*) - (1 - \alpha)(p_{NT} - p_T) \tag{11}$$

If the law of one price holds in the tradable sector ( $p_T = e + p_T^*$) the first RHS term equals zero. Next, substituting (7) into (11) will produce the Balassa Samuelson effect or the external transmission mechanism[2].

$$q = (1 - \alpha^*)\left(\frac{\delta^*}{\gamma^*}(a_T^* - a_{NT}^*)\right) - (1 - \alpha)\left(\frac{\delta}{\gamma}(a_T - a_{NT})\right) \tag{12}$$

$$p - p^* = e + (1 - \alpha)\left(\frac{\delta}{\gamma}(a_T - a_{NT})\right) - (1 - \alpha^*)\left(\frac{\delta^*}{\gamma^*}(a_T^* - a_{NT}^*)\right) \tag{12'}$$

[2] Assuming identical labor intensities for the tradable and non-tradable sector ($\delta = \delta^*, \gamma = \gamma^*$)

| Table 2: Estimates of the Balassa Samuelson effect in previous literature (Chronological order) | | | | | | |
|---|---|---|---|---|---|---|
| **Author(s) + Year of publication** | **Country/Countries + time frame** | **Classification of tradable and non-tradable sector** | **Productivity measure** | **Demand side factors** | **Method of estimation** | **Estimate (% per annum)** |
| De Gregorio et al. (1994) | 14 OECD countries<br><br>(1970-1985) | Tradable: Exports over Total prodcution > 10%;<br><br>Non-tradables: Rest | TFP Solow residuals | Government spending: real gov expenditures over real gdp;<br><br>Income: log of GDP per capita | Ordinary Least Squares: Panel regression | **0.2-0.4**<br>(different controls) |
| De Gregorio & Wolf (1994) | 14 OECD countries<br><br>(1970-1985) | Tradable: Exports over Total prodcution > 10%;<br><br>Non-tradables: Rest | TFP Solow residuals | Terms of trade; Government spending: real gov expenditures over real gdp;<br><br>Income: log of GDP per capita | Ordinary Least Squares: Panel regression | **0.1-0.3**<br>(different controls) |
| Chinn & Johnston (1996) | 14 OECD countries<br><br>(1970-1991) | Tradable: Manufacturing, Mining, Transportation and Agriculture;<br><br>Non-tradable: Services | TFP Solow residuals | Government spending: log of the ratio real gov consumption over real GDP;<br><br>Preferences towards services: GDP per capita | Non-linear Least Squares Time series;<br><br>SUR (Seemingly Unrelated Regression) estimation | **No robust estimates** Time series;<br><br>**0.2-0.9**<br>SUR estimates (different demand side |
| Chinn (1997) | 14 OECD countries<br><br>(1970-1991) | Tradable: Manufacturing, Mining, Transportation and Agriculture;<br><br>Non-tradable: Services | TFP Solow residuals | Government spending: log of the ratio real gov consumption over real GDP;<br><br>Preferences towards services: GDP per capita | Non-linear Least Squares Time series;<br><br>Panel Cointegration | **No robust estimates:** Time series;<br><br>**0.5**<br>Panel cointegration |
| Rother (2000) | Slovenia<br><br>(1993-1998) | Tradable: Manufacturing;<br><br>Non-tradable: Rest;<br><br>Excluded: Agriculture | Labor productivity | Monetary Policy: change in monetary base;<br><br>Fiscal policy: fiscal expenditure over GDP; | Ordinary Least Squares: Time series | **1.2** |
| Halpern & Wyplosz (2001) | 11 CEE countries + Russia<br><br>(1991-1999) | Tradable: Industry;<br><br>Non-tradable: Services | Labor productivity | Income: GDP per capita | General Least Squares: Panel regression | **0.24** |
| De Broeck & Sløk (2001) | 25 transition economies<br><br>EU accession countries vs other transition countries | Tradable: Industry and Construction;<br><br>Non-tradable: services;<br><br>Excluded: Agriculture | Labor productivity | Monetary Policy: Broad money to GDP ratio;<br><br>Fiscal Policy: Government balance;<br><br>Openness of the economy & Terms of | Ordinary Least Squares: Panel regression;<br><br>Pooled Mean Group estimator | **0.8-0.9**<br>Panel OLS<br><br>**0.3-0.6**<br>Pooled Estimator |

| | | | | | | |
|---|---|---|---|---|---|---|
| Coricelli & Jazbec (2001) | 19 transition economies<br><br>(1990-1998) | Tradable: Manufacturing, Mining, Construction, Utilities; Non-tradable: Rest | Labor productivity | Preferences towards services: Share of services consumption over total consumption;<br><br>Government spending: government | Ordinary Least Squares: Panel regression | **0.9** |
| MacDonald & Ricci (2001) | Belgium, Denmark, Finland,<br>France, Italy, Japan, Norway, Sweden, West-Germany, United States<br><br>(1970-1992) | Distribution sector separately;<br><br>Tradable: Manufacturing, Mining, Transportation and Agriculture;<br><br>Non-Tradable: Utilities, Construction and Services | TFP Solow residuals | Importance distribution sector: relative efficiency distribution sector;<br><br>Relative size of net foreign assets & Relative real interest rate;<br><br>Government spending: real gov expenditures over real GDP | Dynamic Ordinary Least Squares: Panel regression | **0.7-0.9**<br>BS effect<br>(different controls);<br><br>**0.4**<br>Importance Dist. sector: |
| Égert<br>(2002) | 6 CEE countries<br><br>(1991-2000) | Tradable: Industry (excl. Construction) and Agriculture;<br><br>Non-tradable: Rest | Labor productivity | None | Ordinary Least Squares: Time series Cointegration | **0.4-3.0**<br>(different countries) |
| Mihaljek & Klau (2003) | 9 CEE countries<br><br>(1995-2001) | Tradable: Manufacturing, Mining, Transportation & Communication;<br><br>Non-tradable: Utilities, Construction, Wholesale & Retail, Financial services, Education & Health;<br><br>Excluded: Agriculture and public administration | Labor productivity | None | Ordinary Least Squares: Time series | **0.2-1.8**<br>(range of BS effect in CEE countries) |
| Égert et al.<br>(2003) | 9 CEE countries<br><br>(1995-2000) | Tradable: Industry (excl. Construction) and Agriculture;<br><br>Non-tradable: Rest | Labor productivity | None | Panel cointegration | **0.3-1.1**<br>(different measures of relative price of non-tradables) |
| Mihaljek & Klau (2008) | 9 CEE countries<br><br>(1995-2007) | Tradable: Manufacturing, Mining and Agriculture;<br><br>Non-tradable: Utilities, Construction, Wholesale & Retail, Financial services, Transportation & Communication;<br><br>Excluded: Non-market services | Labor productivity | None | Ordinary Least Squares: Time series | **0.08**<br>(avg. in the short run)<br><br>**1.1**<br>(avg. in the long run) |

### 2.1.2 Empirics Balassa Samuelson hypothesis

Empirical literature on the Balassa Samuelson effect has flourished ever since the economic rise of Eastern-European and Eastern-Asian countries in the 1990's and early 2000's. Table 2 presents a selection of estimates of the Balassa Samuelson effect which provides a background to the upcoming discussion points[3].

**Overview**

The first thing to note on the empirics is the fact that most literature uses a BEER approach or a simple Balassa Samuelson model, similar to eq. (12), to determine the real exchange rate. The focus is laid on the fundamentals behind the real exchange rate. Largely due to the samples being of countries that underwent trade reforms or economic reform where fundamentals are believed to be the main determinant of the real exchange rate.

The method of estimation is varied among studies but most studies use OLS (panel or time series) or implement panel cointegration methods. Panel studies as a whole were especially prevalent as time series estimations suffered from short time frames. Pooling data into a panel of countries solved this issue to some degree. Chinn (1997) provides a good example where the author estimates eq. (12) with time series and a panel cointegration method. The author was not able find cointegration in levels and, more surprisingly, first differences for the time series estimates for most of the 14 OECD countries. Pooling the data into a panel resulted in a significant panel cointegration estimate. Although, I have to note that the OECD sample is suboptimal to detect a Balassa Samuelson effect in a time series setting. Mihaljek and Klau (2003) and Rother (2000) were able to find cointegration in first differences for a sample of CEE countries which are more suited to detect a Balassa Samuelson effect.

Égert et al. (2006) mention that these panel studies often relied on the assumption of long run homogeneity to interpret the coefficients economically. If the real exchange rate responds similarly to changes in fundamentals across countries, we can assume long run homogeneity. The strength of this assumption depends on the sample of countries, larger samples of countries often having more difficulty to safely assume long run homogeneity (Coricelli and Jazbec (2001), De Broeck and Sløk (2001)). Smaller samples of adjacent countries (Central Eastern European (CEE) countries) might still suffer from a weakened of assumption of long run homogeneity. For example, CEE countries operated different exchange

[3] I follow a similar table layout as used in Mihaljek & Klau (2003)

rate regimes which alters the way fundamentals influence the real exchange rate. This violates the assumption of long run homogeneity to some extent.

A first look at the Balassa Samuelson estimates show a prevalent misalignment between the predicted outcomes of the theoretical model and the empirical estimates. According to the assumptions of the model, the relative price of non-tradable goods is fully determined by the supply side of the economy. Therefore changes in productivity growth should, ceteris paribus, result in one-to-one changes in the real exchange rate. Empirical estimates on the Balassa Samuelson effect generally show values lower than 1 implying an incomplete transmission. This led researchers to look for factors influencing the presumed underestimation of the Balassa Samuelson effect. The following section lays out the reasoning discussed in previous literature.

**Discussion**

It was previously noted in the theoretical model that perfect capital and labor mobility within sectors implies that the Balassa Samuelson effect determines relative price movements completely. Hence, demand side factors do not affect the relative price of non-tradable goods. Therefore a plethora of studies used the simple Balassa Samuelson framework without considering demand side fundamentals. Chinn, Johnston (1996) and Rother (2000) argue that the supply sided nature of the Balassa Samuelson hypothesis can solely explain variation in price movements in the long run. Perfect capital and labor mobility is a fair assumption in the

Table 3: Estimates of Balassa Samuelson effect in the short and long run by Mihaljek & Klau (2008)

$$log\left(\frac{CPI}{CPI^*}\right)_t = c_1 + \beta_0 * log\left(\frac{CPI}{CPI^*}\right)_{t-1} + \beta_1 * log\left(\frac{E_t}{E_{t-1}}\right) + \beta_2 * [(1-\alpha)\, log\left(\frac{LP^T}{LP^{NT}}\right)_t - (1-\alpha^*)\, log\left(\frac{LP^{T*}}{LP^{NT*}}\right)_t] + \varepsilon_t$$

| | Bul. | Cro. | Cz. R. | Est. | Hun. | Lat. | Lit. | Pol. | Rom. | Slk. | Slo. | **Avg.** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Short run* | 0.006 | 0.013 | 0.038 | 0.035 | 0.122 | 0.115 | 0.170 | 0.091 | 0.018 | 0.327 | 0.220 | **0.105** |
| Long run* | 0.031 | 0.165 | 0.169 | 0.947 | 1.549 | 0.619 | 4.628 | 0.903 | 0.441 | 1.961 | 1.686 | **1.191** |

Source: Table 3 in Mihaljek & Klau (2008), estimated coefficients of $\boldsymbol{\beta_2}$ in the short and long run

* All estimated coefficients are significant at the minimum level of 5%

Countries included in the sample: Bulgaria, Croatia, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Romania, Slovakia and Slovenia over mid 1990’s until 2008 long run, but a debatable assumption in the short run. Mihaljek and Klau (2008) give a good representation of the strength of the Balassa Samuelson effect in the short and long run in table 3. It is clearly visible that a large disparity is present between the short and long run strength of the Balassa Samuelson effect. Based on these estimates it seems that perfect capital and labor mobility does not hold in the short run. Therefore price movements in the short run can be explained by market forces such as consumption growth. De Gregorio et al. (1994) add onto this discussion by arguing that not only imperfect capital and labor mobility will allow demand side factors to alter the real exchange rate. Imperfect competition or a violation of PPP for traded goods also allow demand side factors to have an impact. This is especially prevalent in the short run. The authors test for demand side factors by including government expenditures, terms of trade effects and income growth. They note that income growth proxies shifting preferences of consumers towards services as income increases as proposed by the Baumol-Bowen effect. The results are reported in Table 4. It is clearly visible that demand side factors play an important role, additionally the Balassa Samuelson effect is quite small compared to estimates of other studies.

Table 4: Estimates of relevance demand side factors

$$\log P_{i,t} = \beta_1\theta_{i,t} + \beta_2 g_{i,t} + \beta_3 log\ y_{i,t} + \beta_4\Delta\pi_{i,t}$$

| Specification | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_4$ | Obs. |
|---|---|---|---|---|---|
| 1 | 0.234* | 1.974* | 0.281* | | 210 |
| 2 | 0.234* | 1.846* | 0.272* | -0.045* | 210 |

Source: De Gregorio et al. (1994), Table 7.
Sample: 14 OECD countries over 1970-1985
$P_{i,t}$ relative price of non-tradable goods, $\theta_{i,t}$ productivity differential
Demand side variables: $g_{i,t}$ government expenditure over GDP, $y_{i,t}$ GDP per capita (proxy for income level), $\Delta\pi_{i,t}$ first difference of inflation.
* All coefficients represent highly significant estimates (1% level)

According to the theoretical model where the production functions produce output in terms of goods, defining commodities into tradables and non-tradables is critical. The issue becomes apparent considering that the share of tradable and non-tradable goods can be defined accurately for CPI but not for the productivity measure. Creating an accurate division between tradable and non-tradable for the productivity measure is near impossible mainly due to data limitations. Data is often limited to aggregated sectoral data which allows researchers to only divide tradable and non-tradable goods into a tradable and non-tradable sector based on assumptions. Therefore creating a unwanted disparity in the share of tradable and non-tradable goods for the CPI and productivity measure.

The classification of the tradable and non-tradable sector has important implications for the interpretation of the estimates produced. Égert et al. (2003) mentions that the disparity in classification of the sectors for the CPI (component of the dependent variable) and the productivity differentials (independent variable of interest). He states that the magnitude of the Balassa Samuelson effect is partly influenced by the composition of the CPI basket. If the share of non-tradable goods in the basket is relatively low, increases in the relative price of non-tradable goods (due to productivity differences) will not impact CPI by much. Hence, the Balassa Samuelson effect will be small in magnitude as can be seen in table 2 (0.3-0.5).

Égert et al. (2006) expands on the last point by stating that the violation of PPP in the tradable sector is an important factor in explaining the relatively small impact of the Balassa Samuelson effect. The PPI based exchange rate[4] has appreciated which dampens the magnitude of the Balassa Samuelson effect on the CPI based exchange rate. This can be visualized by looking at eq. (11), violation of PPP in tradable sector would make the first RHS term remain in eq. (12) and (12'). Additionally, if the share of non-tradable goods is small in the CPI basket, then the Balassa Samuelson effect would be relatively small.

De Gregorio et al. (1994) also discussed the implications of different exchange rate regimes on the relative price of non-tradables (i.e. internal transmission mechanism). Using a 14 country OECD sample over the 1970-1985, they found that variation in the relative price of non-tradables is significantly lower for core EMS countries (Belgium, Denmark, West German, France and the Netherlands) than other OECD countries. Furthermore, correlations of relative price movements show high correlation values for core EMS economies and lower values for the other OECD countries. The authors argue that this indicates some degree of covariance of prices within quasi-fixed exchange rate regimes (core EMS). This result is not prevalent in the other OECD countries who operate a flexible exchange rate. This discussion can be related back to the problematic nature of the short time frame cross country panel studies where some degree of long run homogeneity is assumed. Differences in exchange rate regimes among your cross country sample introduces an extra element of heterogeneity.

[4] The PPI involves prices of goods produced by industries which are often classified as tradables

## 2.2 Productivity

Literature on the Balassa Samuelson hypothesis does not put much weight on the discussion of the productivity measure used in their estimation. Researchers implement relatively simple measures of productivity such as labor productivity and Solow residuals. Econometric issues involving these productivity measures are often recognized but not acknowledged as a serious threat to reliable estimates. Furthermore, it is worth mentioning that data is mostly insufficient for transition economies to reliably estimate Total Factor Productivity (TFP) for sectors or firms. Ultimately limiting researchers to implement productivity measures such as labor productivity. The following section discusses the issues related to estimation using labor productivity and Solow residuals.

### 2.2.1 Labor productivity

The production functions of the Balassa Samuelson framework suggests the usage of all factors of production in calculating productivity differences among the tradable and non-tradable sector[5]. Hence, the optimal measure for productivity in the tradable and non-tradable sector would be total factor productivity.

Bearing in mind the fact that data is limited, it is logical that researchers resort to labor productivity as a proxy for total factor productivity. Labor productivity simply being the ratio of value added over the number of employees. It becomes quickly apparent why labor productivity is problematic as a measure for productivity. Other factors of production like capital are not considered which enhance output by definition. It is often the case in transition economies that the tradable sector is characterized by being more capital intensive than its non-tradable counterpart. The bias comes into effect when an increase in capital resources is considered. Considering eq. (1) and eq. (2) with the revised assumption that the non-tradable sector is more labor intensive ( $\delta > \gamma$). Given an equal increase in capital resources, it can be observed that output growth will be higher in the tradable sector. Hence, labor productivity differences overestimate actual total productivity differences. Additionally, layoffs are quite problematic in the sense that it increases labor productivity disproportionally compared to total factor productivity.

[5](1) $Y_T = A_T L_T^{\gamma} K_T^{1-\gamma}$, (2) $Y_{NT} = A_{NT} L_{NT}^{\delta} K_{NT}^{1-\delta}$

### 2.2.2 Total factor productivity

There is a vast amount of studies on total factor productivity and how it can be estimated reliably using different methods. One of the main issues with traditional TFP estimations is the endogeneity of input choice, otherwise known as the simultaneity bias. I will discuss the most prominent studies that tried to tackle the simultaneity bias in particular.

Assuming production is given by the Cobb-Douglas production function, we obtain the following:

$$Y_{i,t} = A_{i,t} L_{i,t}^{\beta_l} K_{i,t}^{\beta_k} \tag{13}$$

where $Y_{i,t}$ represent output for firm $i$ at time $t$. $L_{i,t}, K_{i,t}$ represent labor and capital input for firm $i$ at time $t$, respectively. $A_{i,t}$ is the productivity level of firm $i$ at time $t$. Constant returns to scale are considered, therefore $\beta_l + \beta_k = 1$.

Taking logs of eq. (13) I end up with a linear production function:

$$y_{i,t} = \beta_l l_{i,t} + \beta_k k_{i,t} + a_{i,t} \tag{14}$$

where $a_{i,t}$ consists of the firm's TFP $\omega_{i,t}$ and an unobserved error term $\eta_{i,t}$. The firm does not know $\eta_{i,t}$ as it contains exogenous shocks or measurement errors. Firm $i$ knows its productivity level but does not typically share this information. Therefore the researcher does not know either of the terms.

$$y_{i,t} = \beta_l l_{i,t} + \beta_k k_{i,t} + \omega_{i,t} + \eta_{i,t} \tag{14'}$$

Total factor productivity, $\omega_{i,t}$, is traditionally estimated by calculating Solow residuals as shown below.

$$\hat{a}_{i,t} = \widehat{\omega}_{i,t} + \eta_{i,t} = y_{i,t} - \hat{\beta}_l l_{i,t} - \hat{\beta}_k k_{i,t} \tag{15}$$

In essence Solow residuals calculate the residual from the subtraction of actual output minus predicted output based on factor inputs. Keep the error term $\eta_{i,t}$ in mind as it is highly likely to be non-zero. Intuitively, highly productive firms produce more actual output from of a given allocation of factors than low productive firms. Highly productive firms are more efficient in their use of factors of production.

**Issues with estimating eq. (14') and eq. (15)**

Eq. (14') requires that input choices are determined exogenously, independent of a firm's productivity level. To put it more concretely, input choices should not be a function of unobserved determinants such as firm productivity (remember firm productivity being unobservable to the researcher).

Estimating eq. (14') by means of simple OLS introduces biased estimates according to Olley and Pakes (1996). They argue that a firm has at least some prior knowledge of its productivity level $\omega_{i,t}$, therefore input choices will be partially determined by the firm's previous knowledge of $\omega_{i,t}$. What follows is that the error term ($a_{i,t}$, includes $\omega_{i,t}$) of eq. (14) will correlate with the regressors of eq. (14). Correlation between the error term and the regressors will render biased estimates. Positive productivity shocks will lead to increased usage of factors of production. Failing to account for this will result in biased coefficients of the production factors. In general it is difficult to assess the direction of the bias considering different production functions, assumptions and shocks.

Relating the issue back to Balassa Samuelson effect. Due to the endogeneity problem $\hat{\beta}_l$ and/or $\hat{\beta}_k$ will be either over- or underestimated. The bias in the coefficients will consequently bias the present estimate $\widehat{\omega}_{i,t}$ in eq. (15). If one was to obtain biased estimates for the productivity in the tradable and non-tradable sector and substitute them into eq. (12). Then the Balassa Samuelson effect would be either over- or underestimated, more on this in section 3.1.

Over the years several solutions to the endogeneity issue were proposed, most notably Olley & Pakes (1996) and Levinsohn & Petrin (2003). Both methods attempt to proxy for unobserved productivity shocks by either incorporating the firm's investment decision or the usage of intermediate inputs such as materials or energy. The Olley-Pakes method of TFP estimation will be discussed further in the methodology section.

Prior to the OP and LP methods, studies on productivity often implemented fixed effects or Instrumental Variable (IV). Both methods turning out to be fairly problematic as its assumptions would often not hold. Fixed effects estimation relies on the assumption that productivity is firm-specific and time-invariant. Effectively transforming $\omega_{i,t}$ from eq. (14') into $\omega_i$ and resolving the endogeneity issue. Assuming productivity to be time-invariant is rarely valid, especially for transition economies where the Balassa Samuelson effect is supposed to be prevalent. IV is a widely used tool to deal with endogeneity issues and works effectively under two assumptions. The second assumption, the exclusion restriction, is often not satisfied due to a lack of data on valid instruments such as input and output prices.

# 3. Methodology

The empirics of the paper considers two stages. In the first stage (Olley-Pakes) TFP estimates are produced for the tradable and non-tradable sector. Given these TFP estimates, I estimate the Balassa Samuelson effect in the second stage. The following section will be largely based on the theoretical models discussed in 2.1.1 and 2.2.2. These models provide a sufficient background to empirical estimation.

## 3.1. First stage: Total Factor Productivity

Firm-level data is solely available for the transition country Slovenia, therefore Olley-Pakes TFP estimates can only be obtained for the tradable and non-tradable sector in Slovenia. Given this limitation, I estimate Solow residuals for the tradable and non-tradable sector in Germany based on the STAN database. Next, I calculate Olley-Pakes TFP estimates for the tradable and non-tradable sector in Slovenia based on the firm-level dataset. Comparing and plugging the Slovenian Olley-Pakes TFP estimates and Solow residuals allows me to analyze the direction and magnitude of the simultaneity bias. Additionally, output, labor and capital input are all log-linearised in the Solow residuals and Olley-Pakes TFP calculations.

**Sectoral Solow residuals**

The first step involves determining which sectors belong to the tradable (T) and non-tradable (NT) sector. The previous section described the criteria where the ratio of total exports over total output is taken. If it exceeds 10% of total output as used in De Gregorio and Wolf (1994), the sector is classified as tradable[6]. I also calculate weights for each individual sector by taking the simple ratio of the sector's output over the total output of the tradable or non-tradable sector[7]. The weights are used to calculate a weighted average of total factor productivity for the tradable and non-tradable sector.

Originating from eq. (14) I begin with the following specification:

$$y_{S,t} = \beta_l l_{S,t} + \beta_k k_{S,t} + a_{S,t} \quad (16)$$

where $y_{St}$ denotes output for sector $S$ at time $t$, $l_{S,t}$ and $k_{S,t}$ denote labor and capital input for sector $S$ at time $t$. $a_{S,t}$ is an unobserved term containing the sector's TFP and a stochastic error term. Regressing output on capital and labor inputs will produce estimates of $\beta_l$ and $\beta_k$ ($\hat{\beta}_l, \hat{\beta}_k$). Given these estimates I calculate the following (Solow) residual which represents the TFP of sector $S$ at time $t$.

$$\widehat{\omega}_{S,t} = y_{S,t} - \hat{\beta}_l l_{S,t} - \hat{\beta}_k k_{S,t} - \eta_{S,t} \quad (17)$$

where $\widehat{\omega}_{S,t}$ represents the estimated TFP for sector $S$ at time $t$. $\widehat{\omega}_{S,t}$ is estimated for every sector. Next I multiply the TFP estimate by its weight in the tradable or non-tradable sector. Summing up all weighted TFP estimates will result in a weighted average of total factor productivity for the tradable and non-tradable sector.

$$TFP^T = \overline{\omega}_t^T = \sum_{S,t=1}^{St} \theta_{S,t}^T \widehat{\omega}_{S,t}^T \quad (18)$$

$$TFP^{NT} = \overline{\omega}_t^{NT} = \sum_{S,t=1}^{St} \theta_{S,t}^{NT} \widehat{\omega}_{S,t}^{NT} \quad (19)$$

[6] $Classification\ of\ sector\ S = \begin{cases} Tradable\ if & \frac{Total\ exports_S}{Total\ output_S} \geq 0.1 \\ Non-Tradable\ if & \frac{Total\ exports_S}{Total\ output_S} < 0.1 \end{cases}$ where $S$ denotes individual sectors

[7] $Weight\ of\ sector\ S\ at\ time\ t = \theta_{S,t} = \begin{cases} \frac{Total\ output_S}{Total\ output_T}\ if\ S\ is\ Tradable \\ \frac{Total\ output_S}{Total\ output_{NT}}\ if\ S\ is\ Non-Tradable \end{cases}$ where $T, NT$ denote the Tradable and Non-Tradable sector

**Olley-Pakes productivity estimation**

As stated in the literature review on TFP, simple Solow residuals are problematic in the sense that the estimates of capital and labor are biased. Olley and Pakes (1996) developed an alternative method incorporating the investment decision in order to proxy for unobserved productivity shocks. The mathematics and deeper intuition behind the Olley-Pakes methodology can be found in the appendix A.5. The matter is quite complicated and is therefore relegated to the appendix. Luckily, Stata provides code that makes implementing Olley-Pakes fairly simple in estimation.

Following the mathematics in appendix A.5 I end up with a similar specification as in eq. (17) . Implementing the Olley-Pakes method of productivity estimation renders consistent estimates of $\beta_l$ and $\beta_k$. Plugging the newly estimated $\hat{\beta}_l$ and $\hat{\beta}_k$into eq. (15) renders a consistent and unbiased estimate of $\omega_{i,t}$.

(20) $$\widehat{\omega}_{i,t} = y_{i,t} - \hat{\beta}_l l_{i,t} - \hat{\beta}_k k_{i,t} - \eta_{i,t}$$

Similar to eq. (18), the Olley-Pakes TFP estimates are weighted by the contribution of the firm's output to total output of the tradable or non-tradable sector[8]. Summing all the weighted the Olley-Pakes TFP's by their classification will result in weighted averages of Olley-Pakes TFP estimates for the tradable and non-tradable sector.

(21) $$TFP^T = \overline{\omega}_t^T = \sum_{i,t=1}^{it} \theta_{i,t}^T \widehat{\omega}_{i,t}^T \,^{9}$$

[8] $Classification\ of\ firm\ i = \begin{cases} Tradable\ if\ \ \frac{Total\ exports_i}{Total\ output_S} \geq 0.1 \\ Non-Tradable\ if\ \ \frac{Total\ exports_i}{Total\ output_S} < 0.1 \end{cases}$ where $i$ denotes individual firms

[9] $Weight\ of\ firm\ i\ at\ time\ t = \theta_{i,t} = \begin{cases} \frac{Total\ output_i}{Total\ output_T}\ if\ i\ is\ Tradable \\ \frac{Total\ output_i}{Total\ output_{NT}}\ if\ i\ is\ Non-Tradable \end{cases}$ where $T, NT$ denote the Tradable and Non-Tradable sector

(22) $$TFP^{NT} = \overline{\omega}_t^{NT} = \sum_{i,t=1}^{it} \theta_{i,t}^{T} \widehat{\omega}_{i,t}^{NT}$$

The severity and direction of the simultaneity bias can now be determined given the two different methods used for TFP estimates in Slovenia. It is, however, difficult to assess whether the eventual TFP estimate will be over or underestimated. If done right, Olley-Pakes and Levinsohn-Petrin methods of productivity estimation typically produce higher capital coefficients and lower labor coefficients (Van Beveren, 2008). So the overall effect on the residual (TFP) is difficult to assess at this moment.

.

## 3.2 Second stage: Balassa Samuelson effect

The TFP estimates calculated in the previous section are residuals that differ wildly based one estimation method. Comparing these residuals one-to-one would be comparing apples to oranges. The TFP estimates are therefore indexed (base year-quarter Q1:1995) to allow for better comparison. Basing of eq. (12) I respecified the log-linearised equation into a specification applicable for estimation. The indexed TFP estimates from the previous section are plugged in and produces the following:

(23) $$q = \lambda + \beta * \left( (\overline{\omega}_t^{T} - \overline{\omega}_t^{NT}) - (\overline{\omega}_t^{T*} - \overline{\omega}_t^{NT*}) \right) + \mu_t$$

where $q$ represents the real exchange rate as defined by PPP. $\overline{\omega}_t^{T} - \overline{\omega}_t^{NT}$ and its foreign counterpart (denoted by *) constitutes the productivity gap between the tradable and non-tradable sector. $\lambda$ and $\mu_t$ are a constant term and a stochastic error term respectively.

The aim of this paper is to estimate the Balassa Samuelson effect, the effect is captured by the term $\beta$. Interpreting $\beta$ at this stage is likely to be premature as it will suffer from some estimation related issues. The base specification is estimated in levels, it is highly likely that eq. (23) is non-stationary. In order to test non-stationarity easily, I implement the Dickey-Fuller test.

I include a first differences specification to account for the possibility of non-stationary time series in eq. (23).

$$\Delta q = \lambda + \beta * \Delta\big((\overline{\omega}_t^{T} - \overline{\omega}_t^{NT}) - (\overline{\omega}_t^{T*} - \overline{\omega}_t^{NT*})\big) + \mu_t \quad (24)$$

The first differences are denoted by $\Delta$ and are simply taken by the difference in value at time $t$ and time $t-1$. Previous academic research often found stationarity in first differences while suffering from non-stationarity in levels. The downside of implementing first differences is the loss of generality, the interpretation of the coefficients is less convincing. If both series are integrated of order I (1) then the OLS estimator of $\beta$ is typically consistent. It allows for interpretation of the coefficient if it is believed to be unbiased.

I test for cointegration, if both series are to be integrated of order I (1), by implementing a Dickey-Fuller test on the residuals of the first difference specification of eq. (24). Depending on the outcome of this test I can conclude whether the residuals are integrated of order I (0). Integration of order I (0) would imply cointegration of the two series. In economic terms this would imply a long run relationship between the real exchange rate and the productivity differentials.

$\beta$ as stated in eq. (24) is still likely to produce biased estimates if, for example, the assumption of perfect labor and capital mobility does not hold in the short run. If this is believed to be true, then $\beta$ will suffer from omitted variable bias. Other factors that influence the real exchange rate will be captured by $\beta$. Consequently, I include relevant control variables believed and tested to be influential.

$$\Delta q = \alpha + \beta * \Delta\big((\overline{\omega}_t^{T} - \overline{\omega}_t^{NT}) - (\overline{\omega}_t^{T*} - \overline{\omega}_t^{NT*})\big) + \gamma * \Delta\phi + \mu_t \quad (25)$$

where $\phi$ is a vector containing the control variables government expenditures and income level. Both variables are expected to bear a positive coefficient as explained in the following data section. Given the use of first differences and additional control variables I expect to obtain a consistent and unbiased estimator for $\beta$.

# 4. Data

A common theme in this paper is the division of the sections into two parts (Balassa Samuelson and TFP), this also holds for the data section. The two stages of estimation requires data from fundamentally different sources which are discussed in the following section.

In essence, the data obtained covers Slovenia and Germany over a time period running from Q1:1995 to Q4:2006. The choice of nations will be discussed further on. The timeframe covered in this paper has been deliberately chosen. Starting from Q1:1995 allows me to analyze the Balassa Samuelson effect more effectively than if analysis would have started earlier. For Slovenia and other Eastern European countries the years running up to mid 1990 were characterized by substantial fundamental changes to their economies. After the fall of the Berlin wall in 1989, capitalism was reintroduced in the former communist planned economies. Prices had to adjust to market forces, labor and capital were being allocated differently (Jazbec, 2002). The timeframe ends in Q4:2006 because the Slovenian Tolar was replaced by the Euro in Q1:2007.

## 4.1. First stage: Total Factor Productivity

In order to estimate eq. (12) I require TFP estimates for the tradable and non-tradable sector for two countries. Preferably a transition country and a developed country. The Balassa Samuelson effect is most prevalent in this case. Relative productivity gains in the tradable sector are considerably higher in a transition country. The tradable sector in a developed country has matured and shows less substantial gains. Hence, I narrowed the analysis down to Slovenia and Germany. Limiting analysis of the Balassa Samuelson effect to one pair of countries is not optimal, however micro data is limited. Olley-Pakes TFP estimation techniques require firm-level data.

The Slovenian statistical office provides such a dataset and is, as far as I know, the only micro dataset available in the Central-Eastern European countries that is fit for Olley-Pakes TFP analysis. Given the data available for Slovenia, it is desirable to pair Slovenia with a country whom is a significant trading partner. Germany is the main destination for Slovenian exports thus providing a trading partner fit for analysis (OECD, 2018). Germany does not, however, provide a firm-level dataset fit for Olley-Pakes estimation techniques. Therefore, sectoral Solow residuals are calculated.

Sectoral Solow residuals can be estimated given sectoral data on output, labor and capital input, and export status. The STructural ANalysis (STAN) database constructed by the OECD provides sectoral data on all the desired variables. The division of sectors is based on the ISIC classification of industries (revision 3), the sample I extracted contained 58 sectors ranging from agriculture to financial services.

Firm-level data on Slovenian firms has been accessed in a safe room located in the office of the Slovenian statistical bureau (SURS). The dataset contains records of balance sheets and income statements taken from a survey on Slovenian firms. Personal information cannot be disclosed, luckily no personal information is vital for estimation. Data is solely available in annual intervals running from 1994 to 2006. The original sample comprised of about half a million observations. The analyzed sample contains about 250.000 to 300.000 observations, after dropping observations based on missing values and inconsecutive observations not suited for panel analysis. Productivity estimates are interpolated for the second stage to match the quarterly intervals.

The tradability of the firms (firm-level data) or sectors (sectoral data) is determined by the percentage of output exported as proposed by De Gregorio and Wolf (1994). If total exports exceed 10% of total output, the sector or firm will be classified as part of the tradable goods sector. If exports do not exceed 10% of output, the sector or firm will be classified as part of the non-tradable goods sector.

### 4.2. Second stage: Balassa Samuelson effect

The dependent variable of interest in eq. (12) is the real exchange rate[10]. It is comprised of three elements: Consumer Price Index of the host and partner country, and the nominal exchange rate (amount of host currency per unit of partner currency).

Quarterly data on CPI for Slovenia and Germany have been provided by their respective central banks, the Bank of Slovenia and the Central Bank of the Federal Republic of Germany. The nominal exchange rate has been defined as the amount of Slovenian Tolar per Euro. The reasoning behind this definition is the introduction of the Euro as the standard (German) exchangeable currency from 1999 onwards thus covering a large part of the timeframe. The Slovenian central bank provides pre-Euro adjusted exchange rates for the Slovenian Tolar-Euro exchange rate. It is based 1:1 on the fixed Euro - EMU member state currency exchange

[10] $Q = \frac{EP^*}{P}$

rates. Monthly averages of the exchange rate have been extracted and extrapolated to quarterly averages.

**Demand side variables**

The theoretical model predicts price movements to be only affected by the supply side given the assumptions of the model. It is, however, unlikely that these assumptions hold in the real world. Therefore I include demand side control variables to the original Balassa Samuelson specification in eq. (12).

Consistent with most literature I introduce government expenditures into the equation. Considering an increase in government consumption, I expect the relative price of non-traded goods to increase as a consequence. Government expenditures tend to favor non-tradable goods such as public services (education, public transportation), therefore increasing the demand for these goods and hence increasing the price. Given an increase in the relative price of non-tradable goods, I expect the real exchange rate to appreciate. Quarterly data on government expenditure has been obtained from the Slovenian statistical office.

The level of income is also often implemented as an additional control variable. The intuition being that as income increases the consumption of non-tradable goods will increase. Non-tradable goods being primarily services tend to be consumed more often as a country becomes more wealthy. Preferences shift towards the consumption of services. Hence, as transition countries become more wealthy I expect the demand for non-tradable goods to increase and therefore the price of non-tradable goods to increase. The level of income has been defined as GDP per capita in this specification. Quarterly data on GDP and population has been made available by the Slovenian statistical office.

# 5. Results

## 5.1 First Stage

In order for the theoretical model to work some assumptions have to be satisfied to some degree. First and foremost the assumption of equal wage development across the two sectors. I estimated the weighted wage development in annual intervals for the open and sheltered sector. Figure 1 displays the logarithmic development of wages. At first sight a disparity is visible indicating a violation of the assumption. However, wages develop similarly over time and even converge after 2000. Complete wage equalization cannot be assumed though a milder form can be assumed as wages develop similar and converge.

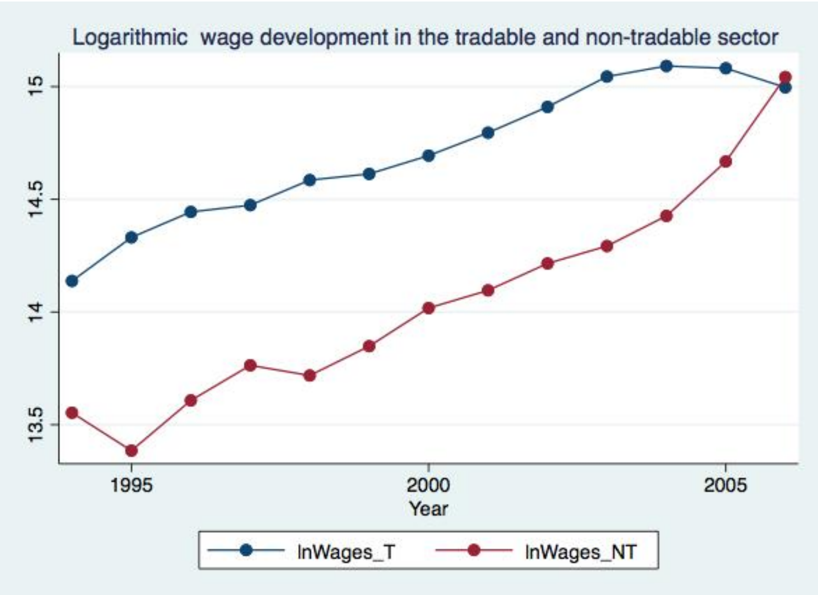



*Figure 1: Weighted logarithmic wage development, Slovenia 1994-2006*

*Source: Annual reports of companies and sole proprietors, SURS, 2018*

Table 5 displays the estimated regression coefficients for the tradable and non-tradable sector in Slovenia. A first glance reveals that the coefficients for labor and capital do not differ much in the two sectors, indicating that labor and capital intensity is similar in both sectors. This result obliges to one of the assumptions of the theoretical model. Furthermore, in almost all cases the coefficients in the Olley-Pakes (OP) and Levinsohn-Petrin (LP) estimates are in line with the predicted biases in section 3.1. The coefficient on labor had an upward bias and has been corrected by the OP and LP methods. Similarly and more importantly, the downward bias on the coefficient of capital has been adjusted.

Table 5: Slovenian firm level regression coefficients

| | | Random Effects | Fixed effects | Olley-Pakes | Levinsohn-Petrin |
|---|---|---|---|---|---|
| Labor | Tradable | 0.722 (.0073) | 0.705 (.0111) | 0.688 (.0038) | 0.691 (.0132) |
| | Non-tradable | 0.692 (.0030) | 0.663 (.0035) | 0.686 (.0032) | 0.711 (.0020) |
| Capital | Tradable | 0.173 (.0046) | 0.152 (.0061) | 0.208 (.0011) | 0.238 (.0154) |
| | Non-tradable | 0.165 (.0019) | 0.141 (.0023) | 0.218 (.0010) | 0.220 (.0007) |

*All coefficients are highly significant, standard errors are reported in brackets.*

*Source: Annual reports of companies and sole proprietors, SURS, 2018*

In order to determine the direction and severity of the simultaneity bias I constructed a graph (see figure 2) visualizing the development of the TFP index based on the two methods of estimation. TFP indices for Solow residual estimations are larger throughout almost the whole timeframe. Therefore indicating an overestimation of total factor productivity if one used Solow residuals for productivity estimation. Looking at the productivity gap it can be stated that the gap is relatively stable for the Olley-Pakes TFP indices. The productivity gap for the Solow residual TFP indices increases rapidly after 2001 after being moderately stable in the first 6 years. Implementing these Solow residual productivity gaps would likely cause overestimation of the Balassa Samuelson effect.

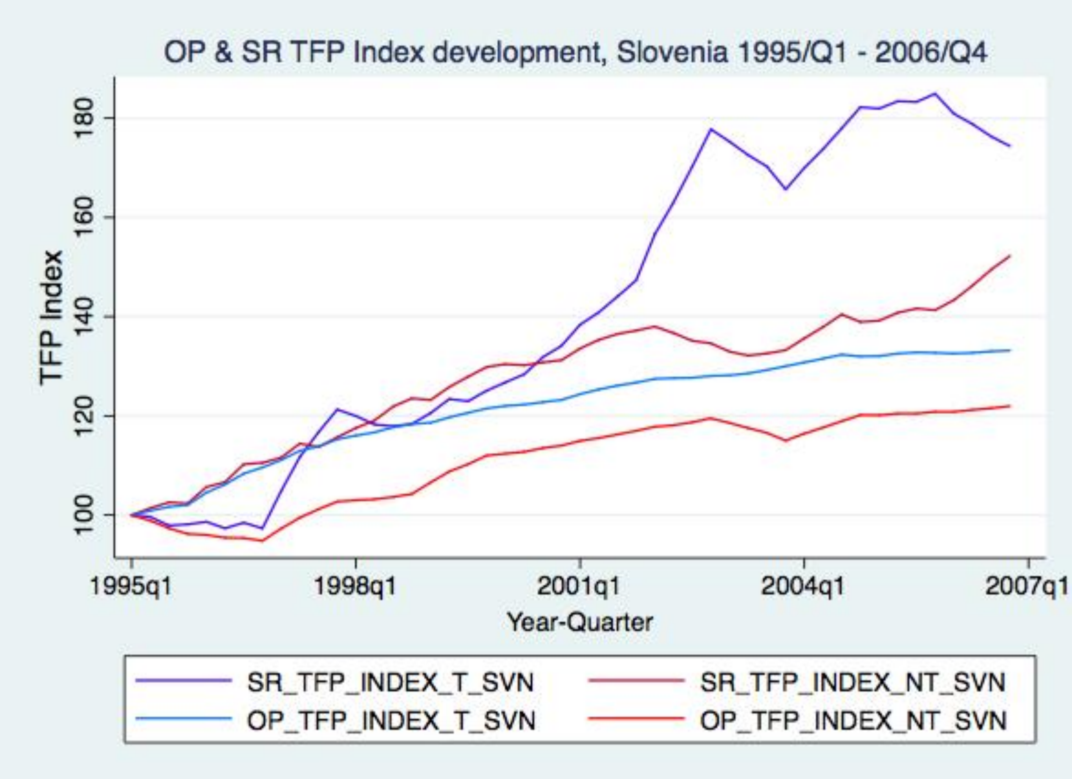



*Figure 2: Olley-Pakes vs. Solow Residual TFP Index development*

*Source: Annual reports of companies and sole proprietors, SURS, 2018*

## 5.2 Second Stage

The following section will lay out the results coming from the three regression specifications discussed in section 4.2. The TFP estimates from the first stage are included in these specifications.

The baseline specification denoted in eq. (23) is simply estimated in levels and no inclusion of any lags. This provides preliminary insight into the Balassa Samuelson effect:

$$q = \lambda + \beta\big((\varpi_t^T - \varpi_t^{NT}) - (\varpi_t^{T*} - \varpi_t^{NT*})\big) + \mu_t$$

| $\beta$ | $\gamma_1$ | $\gamma_2$ | Obs | $R^2$ |
|---|---|---|---|---|
| 0.5561 (0.1390) | - | - | 48 | 0.2581 |

*Table 6: Regression coefficient(s) from baseline specification in eq. (23). Coefficient on the Balassa Samuelson effect is highly significant, standard errors are reported in brackets.*

*Source: Annual reports of companies and sole proprietors, SURS, 2018; OECD, 2018; Banka Slovenije, 2018; Deutsche Bundesbank, 2018*

Interpreting the estimate at face value would state that a 1% increase in the productivity differential difference will increase the real exchange rate by 0.55 %. The estimate for $\beta$ is within the range of estimates of previous literature as visible in table 2. Although less optimistic as the estimates from literature on transition economies where they typically estimated it to be higher. However these studies primarily used labor productivity differentials as opposed to TFP.

The baseline estimate is likely to suffer from non-stationarity as explained in section 3.2. The Dickey-Fuller (see Appendix A.4.2) test performed on both the dependent and independent variable could not reject the null of a unit root in both variables. Drawing conclusions from the baseline estimates should be done with caution as spurious regression might be happening. Applying first differences to the baseline specification produced the following results:

$$\Delta q = \lambda + \beta * \Delta\left((\overline{\omega}_t^T - \overline{\omega}_t^{NT}) - (\overline{\omega}_t^{T*} - \overline{\omega}_t^{NT*})\right) + \mu_t$$

| $\beta$ | $\gamma_1$ | $\gamma_2$ | Obs | $R^2$ |
|---|---|---|---|---|
| 0.0324 (0.0072) | - | - | 47 | 0.3174 |

*Table 7: Regression coefficient(s) from first difference specification in eq. (24). Coefficient on the Balassa Samuelson effect is highly significant, standard errors are reported in brackets.*

*Source: Annual reports of companies and sole proprietors, SURS, 2018; OECD, 2018; Banka Slovenije, 2018; Deutsche Bundesbank, 2018*

At first sight it can be seen that the coefficient has not changed sign, however its magnitude has been reduced significantly to 0.03. Interpreting the coefficient would state that a 1% increase in the growth rate of the productivity differential will lead to a 0.03% increase in the growth rate of the real exchange rate. There is not that much meaning to be taken from this coefficient as it has lost generality and magnitude.

Stationarity has been achieved by first differencing the variables from the baseline specification. I was able to reject the null in the Dickey-Fuller test in both cases. Both variables appear to be integrated of order I (1) (see Appendix A.4.2). I take the residuals from the specification listed above and perform another Dickey-Fuller test. I could not reject the null of a unit root in the residuals. Therefore the residuals are not integrated of order I (0) and no cointegration is present between the first differences of the dependent and independent variable (see Appendix A.4.3). Finding no cointegration in a time series is not unusual. Chinn & Johnston (1996) and Chinn (1997) were also not able to find cointegration in a single time series setting, panel cointegration was present though. Studies on the Balassa Samuelson effect in transition economies did find cointegration in time series. Though one major difference being that these studies used labor productivity which is inherently a suboptimal measure for productivity.

Adding control variables is not going to change much at this point besides testing the robustness of the coefficient on the Balassa Samuelson effect. The controls are directly taken in first differences.

$$\Delta q = \lambda + \beta * \Delta\big((\overline{\omega}_t^T - \overline{\omega}_t^{NT}) - (\overline{\omega}_t^{T*} - \overline{\omega}_t^{NT*})\big) + \gamma_1 * \Delta\, gov + \gamma_2 * \Delta\, gdp + \mu_t$$

| $\beta$ | $\gamma_1$ | $\gamma_2$ | Obs | $R^2$ |
|---|---|---|---|---|
| 0.0290 (0.0064) | -0.0473 (0.0649) | 1.8640 (0.4847) | 46 | 0.4960 |

*Table 8: Regression coefficient(s) from first difference + controls specification in eq. (25). All coefficients highly significant besides an insignificant coefficient on the first difference of government expenditures. Standard errors are reported in brackets.*

*Source: Annual reports of companies and sole proprietors, SURS, 2018; OECD, 2018; Banka Slovenije, 2018;*

There is not much knowledge to gain out of the coefficient on the first differences of government expenditures as it insignificant. The first difference in the GDP growth rate seems to influence the real exchange rate quite a bit, not to say that GDP growth directly influences the real exchange rate. It is more likely that factors closely related to GDP growth influence the real exchange rate directly. The inclusion of controls did not change the coefficient on the Balassa Samuelson effect by much thus showing signs of robustness. Albeit the effect still being rather small.

## 6. Discussion

The main finding of the results shown above is that I was unable to find convincing proof of the Balassa Samuelson effect. Second stage estimates are low in magnitude and no cointegration was found. This is in line with later research in the early 2000's where most studies were sceptic on the importance of the Balassa Samuelson effect. Studies on the Balassa Samuelson effect in transition economies often found convincing results using labor productivity as an approximation for productivity. Using firm-level data to calculate total factor productivity shows that the Balassa Samuelson effect was likely overestimated in those cases. Previous research did mention the downsides of using labor productivity but often did not note that it could affect estimates by much. This paper tries to emphasize the importance of using econometrically sound productivity measures for estimating the Balassa Samuelson effect.

## 6.1 Model-related insights

The results have shown a moderate to small impact of the BS effect similar to more recent literature such as Égert (2002) and Mihaljek and Klau (2007). Plethora of explanations can be derived from the theoretical model and related empirics. In the following section I will cover the most probable causes of the small impact.

### 6.1.1 Balassa Samuelson

The essential assumption of the theoretical model discussed in section 2.1.1 is wage equalization. If not for complete wage equalization, the model would become less powerful. The proportional adjustment of the relative price differential related to the productivity gains in the tradable sector would break down. Complete wage equalization might be unrealistic in the short run. Relative price differentials could be moderate due to incomplete wage equalization. Looking at figure 1 again, it can be observed that wages across the two sectors develop similarly. However, complete wage equalization is not prevalent in this case as wages would have to be identical. Productivity increases in the tradable sector leading to wage increases do not enhance wages identically in the non-tradable sector. Ultimately pressing down the impact of the Balassa Samuelson effect.

The Balassa Samuelson model assumes that PPP holds in the tradable sector by the law of one price. The law of one price nullifies the effect of price adjustments in the tradable sector in eq. (11) leaving us with eq. (12). Importantly, if PPP is believed to not hold in the open sector then eq. (12) becomes invalid and eq. (11) remains. Productivity gains in the tradable sector followed by a surge in wages can then be answered by an increase in the price of tradable goods. This could bring about a downward pressure on the magnitude of Balassa Samuelson effect in eq. (11). In order to roughly assess the validity of the PPP assumption I compare the development of Slovenian CPI and PPI where PPI is a reasonable approximation of the price of tradable goods. Appendix A.1 shows the development of PPI and CPI since Q1:1998, PPI has risen steadily although to a lesser extent. Goods included in PPI represent some share in the calculation of CPI. A rising PPI implies that the rise of CPI is not solely explained by a rise in the price of non-tradable goods. This ultimately weakens the impact of the Balassa Samuelson effect.

Related to the above is the share of non-tradable goods in the calculation of CPI, in the theoretical model visible as $(1 - \alpha)$ in eq. (8). The Balassa Samuelson effect will be small if the share of non-tradable goods is small in the calculation of overall CPI. This can be traced back to eq. (11). The effect of the presumed large relative price differential between the tradable and non-tradable sector will be mitigated if $\alpha$ is large (thus a small share of non-tradables). Additionally, literature on the subject also acknowledges the likelihood of non-tradable goods to be non-market based. An unknown share of these goods are regulated and therefore do not vary naturally to demand and supply conditions. Econometrically often leading to infrequent and large variation which is undesirable.

An often overlooked factor related to transition economies that applied for an EMU membership are the convergence criteria set by the Maastricht treaty. Applicants were required to meet the stringent inflation criterion. In short, obtaining an inflation rate not far above a reference rate compiled of the lowest inflation rate economies in the EMU. The appreciation of the real exchange rate in applicant countries are believed to be caused by the Balassa Samuelson effect where inflation rates are typically high. The inflation criterion basically forbids such high inflation rates. Ultimately forcing Slovenia and others to slow down growth and thus limiting the Balassa Samuelson effect. This is prevalent in the mid 2000's as seen in Appendix A.1 where CPI growth decreased as Slovenia applied for EMU membership.

### 6.1.2 Total factor productivity

First stage estimates of productivity have been corrected for the simultaneity bias by means of the Olley-Pakes estimation methodology. That being said, there are still some concerns regarding their validity. First of all, collinearity issues arise in the estimation of $\beta_l$ in eq. (29') of appendix A.5. It is likely that labor is dependent on the non-parametric term $\emptyset_t(k_{i,t}, i_{i,t})$. It does not vary independently of capital and investment. Collinearity is not necessarily an issue unless variables are highly interdependent. Ackerberg et al. (2006) argue that the latter is the case. Labor is highly if not perfectly collinear with the parametric function of capital and investment. Therefore $\beta_l$ will suffer from large standard errors and imprecise estimates.

Another estimation related issue is the possibility of a present selection bias. The assumption that investment is strictly monotonic in productivity which allows the investment function to be inverted into an productivity function (see appendix A.5). Inversion is not applicable if investment is zero for a given firm or year. These observations are dropped and thus a presumably non-representative sample is left. The selection bias is likely to be present as investment can be characterized by being infrequent and lumpy. However, I do not expect this to affect my estimates much given my definition of investment in eq. (26) in appendix A.5. Here investment is defined in a way where it is unlikely to be negative or zero.

Another source of a selection bias would be the sample used for the Olley-Pakes TFP estimates. In order for Olley-Pakes to work I cannot allow for many missing values or inconsecutive observations. The balanced panel contains a sample which is not entirely similar to the original sample. It mostly comprises medium to large firms who report on all variables needed and also firms who report frequent. Medium to large firms are typically more efficient in production. The consequence being that I have a sample that is likely to produce higher aggregate productivity estimates than if I would have complete data on all firms.

An issue more specific to my research is the widely accepted view that capital was unreliably priced in the post-communist countries in the 1990’s. This would pose a problem for the first stage estimates as it introduces a measurement error. To some extent, I presume this issue to be mitigated in my case. Reasoning being that Slovenia is one of the post-communist countries that adjusted quite well to a market economy. Prices are less diluted by non-market forces present in the communism era in Slovenia. Additionally, Slovenia has put in a good effort to establish a reliable statistics bureau in the early 1990’s. The database used for the TFP calculation grants access to micro data on capital. Data taken from official balance sheets and profit and loss statements. Conventional sectoral data relies on aggregated data which could lead to measurement errors.

## 6.2 Econometric uncertainties

Given my limited data sample for the second stage it is to be expected that econometric issues will arise quite quickly. The obvious issues have been addressed in the methodology and results section. The following will cover more intricate insights related to the estimates presented in the results section.

One major limitation of my research is the unavailability of a firm-level data set for Germany. I could not estimate TFP by means of Olley-Pakes or Levinsohn-Petrin as the dataset used for Germany was limited to sectoral data. The datasets are different and therefore comparing TFP estimates one-to-one in the second stage should be done with some caution.

Time series estimation with about 50 observations is econometrically speaking moderately insufficient. Estimates of small sample time series are more severely affected by outliers and series with a lot of noise over the whole timeframe. Small samples being econometrically insufficient is not necessarily true for all samples. In fact, simple time series estimations in levels and first differences perform perfectly fine for samples with a relatively low amount of noise. The main variables of interest in this setting, real exchange rate and productivity differentials, are relatively stable over the years and do not suffer from a lot of noise.

Interpolation has been applied in some instances to obtain quarterly values. TFP estimates were partially interpolated with the help of quarterly growth rates of added value. Growth rates taken from national accounts for the tradable and non-tradable sector. The residual has been interpolated linearly. Applying interpolation introduces some degree of autocorrelation by definition. In terms of levels estimation, it has to be taken into consideration as autocorrelation will likely be present.

One thing that is often overlooked in the literature is the plausible presence of reverse causality. One could argue that a depreciated nominal exchange rate will encourage a surge in demand for traded goods. The surplus in demand has to be met with an increase in supply which could be accomplished by an increase in productivity. Note that the dependent variable is only partially constructed by the nominal exchange rate. In this case the causality runs the other way around to some degree.

## 7. Conclusion

This paper aimed to estimate the Balassa Samuelson effect based on Olley-Pakes TFP estimates. The results shown favor the view of later Balassa Samuelson literature where the effect is small or non-existent. In this case the small effect can be traced back to smaller productivity differentials. TFP Solow residuals as a measure for productivity exaggerates the productivity difference between the tradable and non-tradable sector. Olley-Pakes TFP differentials were smaller and therefore the Balassa Samuelson effect is weaker. Previous studies often glanced over the issues related to labor productivity and Solow residuals as approximations for productivity. Given the results I attempted to show the importance of using econometrically sound total factor productivity measures.

The major limitation of this paper is the lack of firm-level data for Germany which did not allow me to do Olley-Pakes TFP analysis on both countries involved. What follows is my main recommendation for further research. Firm-level data becomes more (publicly) available and may in the future allow for panel analysis based on Olley-Pakes TFP estimates of multiple countries. This greatly enhances the ability to estimate econometrically sound TFP estimates suited for better second stage estimates.

# 8. References


Ackerberg, D., Caves, K., & Frazer, G. (2006). *Structural identification of production functions*. University Library of Munich, Germany.

Arnold, J. M. (2005). Productivity estimation at the plant level: A practical guide. *Unpublished manuscript*, *27*.

Asea, P., & Corden, W. M. (1994). The Balassa-Samuelson Model: an Overview. Review of International Economics, 2(3), 191-200.

Balassa, B. (1964). The purchasing-power parity doctrine: a reappraisal. Journal of political Economy, 72(6), 584-596.

Banka Slovenije. (2018). *Data series*. Retrieved from https://www.bsi.si/en/statistics/data-series/data-series

Chinn, M., & Johnston, L. (1996). Real Exchange Rate Levels, Productivity and Demand Shocks: Evidence from a Panel of 14 Countries (No. 5709). National Bureau of Economic Research, Inc.

Chinn, M. (1997). Sectoral Productivity, Government Spending and Real Exchange Rates: Empirical Evidence for OECD Countries (No. 6017). National Bureau of Economic Research, Inc.

Coricelli, F., & Jazbec, B. (2001). Real Exchange Rate Dynamics in Transition Economies (No. 2869). CEPR Discussion Papers.

De Broeck, M., & Sløk, T. (2001). Interpreting real exchange rate movements in transition countries. Journal of International Economics, 68(2), 368-383.

De Gregorio, J., & Wolf, H. C. (1994). Terms of trade, productivity, and the real exchange rate (No. w4807). National Bureau of Economic Research.

De Gregorio, J., Giovannini, A., & Wolf, H. C. (1994). International evidence on tradables and nontradables inflation. European Economic Review, 38(6), 1225-1244.

De Loecker, J., & Konings, J. (2006). Job reallocation and productivity growth in a post-socialist economy: Evidence from Slovenian manufacturing. *European Journal of Political Economy*, *22*(2), 388-408.

Deutsche Bundesbank. (2018). *Time series database*. Retrieved from https://www.bundesbank.de/en/statistics/time-series-databases

Égert, B. (2002). Estimating the impact of the Balassa–Samuelson effect on inflation and the real exchange rate during the transition. Economic Systems, 26(1), 1-16.

Égert, B., Drine, I., Lommatzsch, K., & Rault, C. (2003). The Balassa–Samuelson effect in Central and Eastern Europe: myth or reality?. Journal of comparative Economics, 31(3), 552-572.

Égert, B., Halpern, L., & MacDonald, R. (2006). Equilibrium exchange rates in transition economies: taking stock of the issues. Journal of Economic surveys, 20(2), 257-324.

Halpern, L., & Wyplosz, C. (2001). Economic transformation and real exchange rates in the 2000s: the Balassa-Samuelson connection. Economic Survey of Europe, 1, 227-239.

Jazbec, B. (2002). Balassa-Samuelson effect in transition economies: the case of Slovenia.

Levinsohn, J., & Petrin, A. (2003). Estimating production functions using inputs to control for unobservables. *The Review of Economic Studies*, *70*(2), 317-341.

MacDonald, R., & Ricci, L. (2001). PPP and the Balassa Samuelson Effect: the Role of the Distribution Sector (No. 442). CESifo Working Paper.

Mihaljek, D., & Klau, M. (2003). The Balassa–Samuelson effect in central Europe: a disaggregated analysis. Comparative Economic Studies, 46(1), 63-94.

Mihaljek, D., & Klau, M. (2008). Catching-up and inflation in transition economies: the Balassa-Samuelson effect revisited (No. 270). Bank for International Settlements.

OECD. (2018). *Quarterly National Accounts*. Retrieved from https://stats.oecd.org/#

OECD. (2018). *Structural Data Analysis (STAN)*. Retrieved from https://stats.oecd.org/#

Olley, G. S., & Pakes, A. (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica*, *64*(6), 1263-1297.

Rother, C. P. (2000). The impact of productivity differentials on inflation and the real exchange rate: an estimation of the Balassa-Samuelson effect in Slovenia. IMF Country Report. Republic of Slovenia: Selected Issues 00/56, 26-39.

Samuelson, P. A. (1964). Theoretical notes on trade problems. The Review of Economics and Statistics, 145-154.

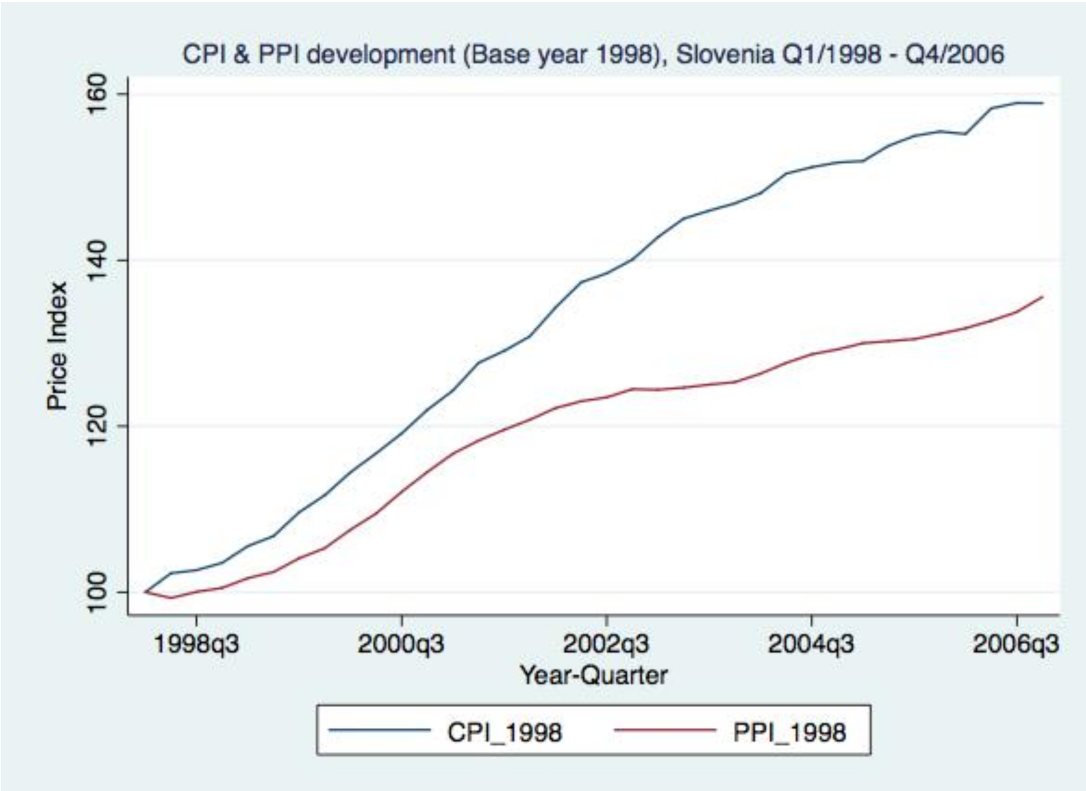
Statistični urad Republike Slovenije (SURS). (2018). *SI-STAT database*. Retrieved from https://pxweb.stat.si/pxweb/dialog/statfile1.asp

Statistični urad Republike Slovenije (SURS). (2018). *Survey: Annual reports of companies and sole proprietors.* Ljubljana. Retrieved from safe room access

Van Beveren, I. (2012). Total factor productivity estimation: A practical review. *Journal of economic surveys*, *26*(1), 98-128.

# Appendix

## A.1 CPI & PPI development




*Source: OECD, 2018; SURS, 2018; Deutsche Bundesbank, 2018*

## A.2 Stata output regression First Stage OP, LP and FE

```
Olley-Pakes tradable sector


levpet lnav if tradable==1, free(lnlabour) proxy(lninv) capital(lnTA) i(ID_customnr) t(year) reps(2)


Levinsohn-Petrin productivity estimator

Dependent variable represents value added.      Number of obs      =     38179
Group variable (i): ID_customnr                 Number of groups   =     38400
Time variable (t): year
                                                Obs per group: min =         1
                                                               avg =       7.6
                                                               max =        13

------------------------------------------------------------------------------
        lnav |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lnlabour |   .6876176   .0038393   179.10   0.000     .6800928    .6951425
        lnTA |   .2077354   .0011217   185.19   0.000     .2055369    .2099339
```

Olley-Pakes non-tradable sector

```
levpet lnav if tradable==0, free(lnlabour) proxy(lninv) capital(lnTA) i(ID_customnr) t(year) reps(2)
```

```
Levinsohn-Petrin productivity estimator

Dependent variable represents value added.      Number of obs      =    236367
Group variable (i): ID_customnr                 Number of groups   =     38400
Time variable (t): year
                                                Obs per group: min =         1
                                                               avg =       7.6
                                                               max =        13

------------------------------------------------------------------------------
        lnav |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lnlabour |   .6864168   .0031476   218.08   0.000     .6802476     .692586
        lnTA |   .2178533    .001041   209.28   0.000     .2158131    .2198936
```

Levinsihn-Petrin tradable sector

```
levpet lnav if tradable==1, free(lnlabour) proxy(lnmaterials) capital(lnTA) i(ID_customnr) t(year) reps(2)
```

```
Levinsohn-Petrin productivity estimator

Dependent variable represents value added.      Number of obs      =     34149
Group variable (i): ID_customnr                 Number of groups   =     38400
Time variable (t): year
                                                Obs per group: min =         1
                                                               avg =       7.6
                                                               max =        13

------------------------------------------------------------------------------
        lnav |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lnlabour |   .6910346   .0132396    52.19   0.000     .6650856    .7169837
        lnTA |   .2382256   .0153901    15.48   0.000     .2080616    .2683896
```

Levinsihn-Petrin non-tradable sector

```
levpet lnav if tradable==0, free(lnlabour) proxy(lnmaterials) capital(lnTA) i(ID_customnr) t(year) reps(2)
```

```
Levinsohn-Petrin productivity estimator

Dependent variable represents value added.      Number of obs      =    208220
Group variable (i): ID_customnr                 Number of groups   =     38400
Time variable (t): year
                                                Obs per group: min =         1
                                                               avg =       7.6
                                                               max =        13

------------------------------------------------------------------------------
        lnav |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lnlabour |   .7106874   .0020035   354.71   0.000     .7067605    .7146143
        lnTA |   .2199211   .0007367   298.53   0.000     .2184772     .221365
```

```
Fixed effects tradable sector


xtreg lnav lnlabour lnTA if tradable==1, fe robust cluster(ID_customnr)

Fixed-effects (within) regression               Number of obs      =     38514
Group variable (i): ID_customnr                 Number of groups   =      9646

R-sq:  within  = 0.6143                         Obs per group: min =         1
       between = 0.8863                                        avg =       4.0
       overall = 0.9140                                        max =        13

                                                F(2,9645)          =   6476.05
corr(u_i, Xb)  = 0.2853                         Prob > F           =    0.0000

                          (Std. Err. adjusted for 9646 clusters in ID_customnr)
------------------------------------------------------------------------------
             |               Robust
        lnav |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lnlabour |   .7054572   .0111237    63.42   0.000     .6836525    .7272619
        lnTA |   .1519934   .0061125    24.87   0.000     .1400116    .1639752


Fixed effects non-tradable sector


xtreg lnav lnlabour lnTA if tradable==0, fe robust cluster(ID_customnr)

Fixed-effects (within) regression               Number of obs      =    240612
Group variable (i): ID_customnr                 Number of groups   =     35488

R-sq:  within  = 0.5637                         Obs per group: min =         1
       between = 0.8476                                        avg =       6.8
       overall = 0.8333                                        max =        13

                                                F(2,35487)         =  38171.79
corr(u_i, Xb)  = 0.4176                         Prob > F           =    0.0000

                         (Std. Err. adjusted for 35488 clusters in ID_customnr)
------------------------------------------------------------------------------
             |               Robust
        lnav |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lnlabour |   .6627096   .0035418   187.11   0.000     .6557676    .6696517
        lnTA |   .1413443   .0022731    62.18   0.000     .1368888    .1457997
```

*Source: Annual reports of companies and sole proprietors, SURS, 2018; OECD, 2018*

## A.4 Stata output regression Second stage

### A.4.1. Regression specifications eq. (27), eq. (28) and eq. (29)

```
reg RER BalSamEff_Index_OP

      Source |       SS           df       MS      Number of obs   =        48
-------------+----------------------------------   F(1, 46)        =     16.00
       Model |  361.202769         1  361.202769   Prob > F        =    0.0002
    Residual |  1038.15097        46  22.5684992   R-squared       =    0.2581
-------------+----------------------------------   Adj R-squared   =    0.2420
       Total |  1399.35373        47  29.7734837   Root MSE        =    4.7506

------------------------------------------------------------------------------------
               RER |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------------+----------------------------------------------------------------
BalSamEff_Index_OP |   .5561325   .1390126     4.00   0.000     .2763147    .8359502
             _cons |   139.8581   1.522356    91.87   0.000     136.7937    142.9224
------------------------------------------------------------------------------------
```

```
reg d.RER d.BalSamEff_Index_OP

      Source |       SS           df       MS      Number of obs   =        47
-------------+----------------------------------   F(1, 45)        =     20.46
       Model |  .019415041         1  .019415041   Prob > F        =    0.0000
    Residual |  .041761051        45  .000949115   R-squared       =    0.3174
-------------+----------------------------------   Adj R-squared   =    0.3018
       Total |  .061176093        46  .001359469   Root MSE        =    .03081

------------------------------------------------------------------------------
          D.RER |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------+-------------------------------------------------------------
BalSamEff_Index_OP |
            D1. |   .0324353   .0071715     4.52   0.000     .0179822    .0468885
          _cons |   4.907144   .0162547   301.89   0.000     4.874385    4.939903
------------------------------------------------------------------------------

reg d.RER d.BalSamEff_Index_OP d.lnGov d.lnGDP

      Source |       SS           df       MS      Number of obs   =        46
-------------+----------------------------------   F(3, 42)        =     13.78
       Model |  .030343145         3  .010114382   Prob > F        =    0.0000
    Residual |  .030832947        42  .000734118   R-squared       =    0.4960
-------------+----------------------------------   Adj R-squared   =    0.4600
       Total |  .061176093        45  .001359469   Root MSE        =    .02709

------------------------------------------------------------------------------
                RER |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
--------------------+---------------------------------------------------------
 BalSamEff_Index_OP |
                D1. |   .0288889   .0064057     4.51   0.000     .0159617     .041816
                    |
              lnGov |
                D1. |  -.0473366    .064901    -0.73   0.470    -.1783122     .083639
                    |
              lnGDP |
                D1. |   1.864017   .4847421     3.85   0.000      .885768    2.842266
                    |
              _cons |   4.868714   .0174804   278.52   0.000     4.833437    4.903991
------------------------------------------------------------------------------
```

*Source: Annual reports of companies and sole proprietors, SURS, 2018; Banka Slovenije, 2018; Deutsche Bundesbank, 2018*

## A.4.2 Dickey-Fuller tests on levels and first differences

```
dfuller RER, lags(4) trend regress

Augmented Dickey-Fuller test for unit root         Number of obs   =        43

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -2.641            -4.214            -3.528            -3.197
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.2612

------------------------------------------------------------------------------
D.RER        |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         RER |
         L1. |  -.3290961   .1245925    -2.64   0.012    -.5817814   -.0764108
         LD. |   .1464212    .144335     1.01   0.317    -.1463037    .4391462
        L2D. |  -.1685881   .1435626    -1.17   0.248    -.4597465    .1225703
        L3D. |   .3046333   .1294951     2.35   0.024     .0420051    .5672615
        L4D. |   .2646003   .1233702     2.14   0.039     .0143939    .5148067
      _trend |  -.0008449   .0003502    -2.41   0.021    -.0015552   -.0001346
       _cons |   1.657893   .6291343     2.64   0.012      .381949    2.933836
------------------------------------------------------------------------------
```

```
dfuller BalSamEff_Index_OP, lags(4) regress

Augmented Dickey-Fuller test for unit root         Number of obs   =        43

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -1.259            -3.628            -2.950            -2.608
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.6479

------------------------------------------------------------------------------------
D.                 |
BalSamEff_Index_OP |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------------+----------------------------------------------------------------
BalSamEff_Index_OP |
               L1. |  -.1637481   .1301103    -1.26   0.216    -.4273767    .0998805
               LD. |  -.1110868   .1939542    -0.57   0.570    -.5040753    .2819016
              L2D. |  -.0089032   .1818741    -0.05   0.961    -.3774152    .3596088
              L3D. |  -.0663598   .1758591    -0.38   0.708    -.4226843    .2899646
              L4D. |   .1663522   .1638383     1.02   0.317    -.1656158    .4983202
                   |
             _cons |   1.591388   1.462404     1.09   0.284    -1.371724    4.554499
------------------------------------------------------------------------------------


dfuller d.RER, lags(4) regress

Augmented Dickey-Fuller test for unit root         Number of obs   =        42

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -3.678            -3.634            -2.952            -2.610
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0044

------------------------------------------------------------------------------
      D2.RER |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         RER |
         LD. |  -1.290119   .3507806    -3.68   0.001    -2.001535   -.5787026
        LD2. |    .236036   .3107814     0.76   0.453    -.3942579    .8663299
       L2D2. |  -.1099344    .257722    -0.43   0.672    -.6326188      .41275
       L3D2. |   .0123707    .196082     0.06   0.950    -.3853019    .4100434
       L4D2. |   .1933914   .1254603     1.54   0.132    -.0610538    .4478366
             |
       _cons |   -.003394   .0014226    -2.39   0.022    -.0062792   -.0005087

dfuller d.BalSamEff_Index_OP, lags(4) regress

Augmented Dickey-Fuller test for unit root         Number of obs   =        42

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -3.693            -3.634            -2.952            -2.610
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0042

------------------------------------------------------------------------------------
D2.                |
BalSamEff_Index_OP |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------------+----------------------------------------------------------------
BalSamEff_Index_OP |
               LD. |  -1.694167   .4588089    -3.69   0.001    -2.624674   -.7636591
              LD2. |   .5004833   .4079601     1.23   0.228    -.3268982    1.327865
             L2D2. |   .3246411   .3226528     1.01   0.321    -.3297292    .9790113
             L3D2. |    .092833   .2448184     0.38   0.707    -.4036818    .5893478
             L4D2. |    .110095     .15808     0.70   0.491    -.2105061    .4306961
                   |
             _cons |  -.2801448   .4769339    -0.59   0.561    -1.247412     .687122
```

*Source: Annual reports of companies and sole proprietors, SURS, 2018; Banka Slovenije, 2018; Deutsche Bundesbank, 2018*

### A.4.3 Dickey-Fuller test on residuals of eq. (28) to test for cointegration

```
dfuller residual_firstdiff, lags(4) regress

Augmented Dickey-Fuller test for unit root         Number of obs   =        38

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -1.865            -3.662            -2.964            -2.614
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.3489

------------------------------------------------------------------------------------
D.                  |
residual_firstdiff |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------------+----------------------------------------------------------------
residual_firstdiff |
                L1. |  -.1555675   .0834361    -1.86   0.071    -.3255213    .0143863
                LD. |  -.2779758   .1697632    -1.64   0.111    -.6237721    .0678205
               L2D. |   -.309859   .1613103    -1.92   0.064    -.6384373    .0187194
               L3D. |  -.1326354   .1580571    -0.84   0.408    -.4545871    .1893164
               L4D. |   .0431129   .1469789     0.29   0.771    -.2562734    .3424991
                    |
              _cons |  -.0040079   .0024187    -1.66   0.107    -.0089347    .0009188
```

*Source: Annual reports of companies and sole proprietors, SURS, 2018; Banka Slovenije, 2018; Deutsche Bundesbank, 2018*

## A.5 Olley-Pakes Mathematics

Olley & Pakes start out with a similar Cobb-Douglas production function described in eq. (13) and the log-linearized eq. (14). It is assumed that capital is a dynamic input whereas labor is a non-dynamic input. Capital inputs at time $t$ are determined at time $t-1$ by means of investment. Investment involves capital accumulation which introduces the dynamic element to the input of capital.

(26) $$k_{i,t} = (1-\kappa)k_{i,t-1} + i_{i,t-1}$$

where capital of firm $i$ at time $t$ is a function of the depreciated capital and investment at time $t-1$. One can intuitively see that the simultaneity issue for $k_{i,t}$ is solved as it is determined at time $t-1$ while $\omega_{i,t}$ is determined at time $t$. Labor still suffers from endogeneity issues as it is determined at time $t$ just as $\omega_{i,t}$. The following will attempt to tackle the issue.

Olley and Pakes define investment of firm $i$ at time $t$ as a function of productivity and capital at time $t$. Labor is not included into the investment function as it is a non-dynamic input. Additionally, investment does not depend on historical values of $\omega_{i,t}$ (and therefore labor) due to the first order Markov assumption[11].

(27) $$i_{i,t} = f_t(\omega_{i,t}, k_{i,t})$$

OP assume that investment is strictly increasing in productivity ($\omega_{i,t}$), hence productivity is strictly monotonic. Assuming investment to be strictly monotonic in productivity allows productivity to become an inverted function of investment and capital at time $t$. Productivity is now described in terms of observables which is crucial.

(28) $$\omega_{i,t} = f_t^{-1}(i_{i,t}, k_{i,t})$$

$f_t$ is indexed by a time component as OP indicate that the function is partly determined by demand conditions, input prices et cetera. The solution to the function becomes very difficult to obtain, therefore it is treated non-parametrically.
Given eq. (28) we can rewrite eq. (14').

(29) $$y_{i,t} = \beta_l l_{i,t} + \beta_k k_{i,t} + f_t^{-1}(i_{i,t}, k_{i,t}) + \eta_{i,t}$$

(29') $$y_{i,t} = \beta_l l_{i,t} + \emptyset_t(k_{i,t}, i_{i,t}) + \eta_{i,t}$$ [12]

where $\emptyset_t$ is approximated by a third or fourth order polynomial of investment and capital, also known as the efficiency level[13]. $\eta_{i,t}$ is now the only unobservable term. Performing a regression on eq. (29') will render a consistent estimate of $\beta_l$. I now have an unbiased and consistent estimate for the labor coefficient ($\widehat{\beta_l}$), the estimate can later be used in the calculation of the Solow residual. $\beta_k$ cannot be estimated right away because of high collinearity between $k_{it}$

[11] $P(w_{i,t+1}|w_{i,t}, w_{i,t-1}, w_{i,t-2}, \dots) = P(w_{i,t+1}|w_{i,t})$, knowledge of productivity at t has just as much predictive power as knowledge of all previous productivity levels (including productivity at t). It is reasonable to assume that a firm forms expectations on its future productivity based solely on its most recently known productivity level.

[12] $\emptyset_t = k_{i,t} + i_{i,t} + {k_{i,t}}^2 + {i_{i,t}}^2 + {k_{i,t}}^3 + {i_{i,t}}^3 + k_{i,t} * {i_{i,t}}^2 + {k_{i,t}}^2 * i_{i,t}$

[13] $\emptyset_t(k_{it}, i_{i,t}) = \beta_k k_{it} + f_t^{-1}(i_{i,t}, k_{i,t})$

and $f_t^{-1}(i_{i,t}, k_{i,t})$, additionally it is treated non-parametrically. In order to obtain $\beta_k$, the following equation is used:

$$(30) \qquad y_{i,t} - \widehat{\beta_l} l_{i,t} = \alpha + \beta_k k_{i,t} + \rho\left(\widehat{\emptyset}_{t-1} - \beta_k k_{i,t-1}\right) + \varepsilon_{i,t} + \eta_{i,t}$$

$\varepsilon_{i,t}$ constitutes an unobservable 'news' component related to the efficiency term, importantly $k_{i,t}$ is independent from $\varepsilon_{i,t}$[14]. $\rho$ is a transformation of the estimated efficiency level, while taking into account that capital in time t depends on capital and investment in in time t-1. Given the structure of eq. (26) with present values of capital in the regression specification and lagged values of capital in the polynomial $\rho$, I implement Non-linear Least Squares (NLS) estimation consistent with previous literature. Regressing eq. (30) will result in a reliable estimate of $\beta_k$. Consistent estimates of $\beta_l$ and $\beta_k$ are now obtained. Plugging the newly estimated $\hat{\beta}_l$ and $\hat{\beta}_k$ into eq. (15) renders a consistent and unbiased estimate of $\omega_{i,t}$.

[14] $E\left(\varepsilon_{i,t} \middle| k_{i,t}\right) = 0$, given by the assumption that capital inputs are dynamic and therefore determined in t-1