Estimation of the pace and rate of emigration using SETAR models: Econometric analysis based on data from Poland

This paper, which draws on monthly data from 2002–2011, aims to prove that Polish emigration proceeds in spurts – it is, in other words, characterized by relatively long spells of stability that are followed by sudden outflows triggered by such factors as changes in unemployment or currency exchange rates. In doing so, it first seeks to find out whether SETAR (threshold) models are better-placed than ARMA/ARIMA models to estimate the pace and rate of emigration. It transpires from our econometric analysis that this is indeed the case, which, in turn, allowed us to estimate the threshold values of selected transition variables – emigration, unemployment, purchasing power parity, wages, GDP and the PLN/EUR exchange rate – that set off the outflows. These findings per se merit recognition due to their socio-economic and demographic implications. The study contributes to the literature by proposing an original approach to analyzing the dynamics of migratory processes.

Keywords: emigration, Poland, econometric modelling, SETAR/ARIMA models

JEL classification: J61

Szanowanie tempa i poziomu emigracji przy pomocy modeli SETAR – analiza ekonometryczna na podstawie danych z Polski

Introduction

The accession of former communist countries to the European Union in 2004 triggered an unprecedented wave of migration from the East to the West [Zientara, 2011; Goździak, 2014]. Indeed, the last decades have seen massive outflows of people from less affluent Eastern European economies, such as Poland, Latvia and Lithuania [OECD, 2013]. High unemployment rates and relatively low wages have encouraged hundreds of thousands of Eastern Europeans to seek brighter prospects abroad. And it is Poles who account for the largest proportion of the Eastern European immigrant population in Western Europe. In fact, approximately 2.3 million Polish citizens have left the country for at least three months since 2004 [CSO, 2015]. Most of them decided to move to the UK, where, of late, they have become – alongside other Eastern Europeans (such as Romanians) – the target of xenophobic attacks. In particular, they have been blamed for stealing jobs, driving down wages and abusing national welfare systems. All this, in turn, has prompted multidisciplinary research into the problematics of Polish post-accession migration [White, Ryan, 2008; Ryan et al., 2009; Burrell, 2009; Drinkwater, Eade, Garapich, 2009; White, 2011; McGhee, Heath, Trevena, 2013; Fihel, Grabowska-Lusinska, 2014; Goździak, 2014; Isański, Mleczko, Seredyńska-Abou Eid, 2014; Kilkey, Plomien, Perrons, 2014; Mostowska, 2014].

Much of this research work, drawing on evidence from Britain (and, to a smaller degree, Scandinavia), has explored the patterns of migratory movements and the experiences of Polish immigrants. In this context, it is argued that post-accession migrations, being marked by (frequent) temporary stays and returns, have become ‘much more differentiated’ [Goździak, 2014, p. 1] and ‘much more circular than 20 or 30 years earlier’ [Isański, Mleczko, Seredyńska-Abou Eid, 2014, p. 5]. What lies behind these changes is a combination of specific circumstantial factors, such as availability of cheap flights due to the rise of low-cost airlines and enforcement of (EU-wide) freedom-of-movement rules. Moreover, there is strong evidence that, from a sociological perspective, there has been a noticeable shift in the traditional migratory pattern whereby it was a male breadwinner who moves abroad and then sends remittances back home to support his family [White, 2011]. In fact, more and more married women work abroad, with their husbands and children staying back in Poland [Cieślińska, 2014]. At the same time, a growing number of Poles decide to move to the UK in a bid to realize their potential and/or to escape the intolerance and petty-mindedness of Polish society rather then only
to find better-paid employment [Isański, Mleczko, Seredyńska-Abou Eid, 2014; Kilkey, Plomien, Perrons, 2014; Main, 2014]. All these developments have led to ‘more diverse and floating populations’ [Goździak, 2014, p. 1].

Other studies explore the impact of post-accession immigration on the British economy, with particular emphasis placed on the wages and employment of native workers [House of Lords, 2008; Drinkwater, Eade, Garapich, 2009; Reed, Lattore, 2009; Somerville, Sumption, 2009]. What transpires from this strand of research is that immigrants ‘contribute more to the Treasury in taxes than they take out in benefits and services’ [The Economist, 2013, p. 8]. Crucially, it is increasingly well-documented that the potentially negative effects of immigration – such as depression of the wages of low-skilled Brits or an increase in unemployment among this group – are, by and large, inconsequential and short-lived. However, these findings, coupled with the evidence that British employers rate highly Polish employees’ work ethic, has done little to reverse the public backlash against immigration. The rise and growing popularity of the anti-immigrant and anti-EU UK Independence Party also confirms that view.

These considerations emphasize the continuous need for extensive, multidisciplinary research into post-enlargement Polish migratory processes. Pertinently, the present paper, which draws on monthly/quarterly data from 2002–2011, aims to prove that Polish emigration proceeds in spurts – in other words, it is characterized by relatively long spells of stability that are followed by sudden migratory outflows, which, in turn, are triggered by such factors as changes in unemployment or currency exchange rates. In doing so, the study seeks to find out whether SETAR (threshold) models are better-placed than ARMA/ARIMA models to estimate the pace and rate of emigration. Therefore, it also proposes – on condition that the above assumption is borne out – to estimate the threshold values of selected transition variables (emigration, unemployment, purchasing power parity, wages, GDP growth and the PLN/EUR exchange rate) that set off the outflows. These variables represent – or are good proxies for – typical pull and push factors [Lee, 1966]. As is widely acknowledged, high unemployment and/or low GDP growth and/or low wages at home usually push people out of the country; PPP and exchange rate movements have the same effect by making work abroad more (financially) rewarding while (prior) emigration acts as a pull (since it is easier for those who consider emigrating to decide to move abroad if they can count on relatives or friends who are already in the destination country).

It is true that there are several recent studies focusing on Polish post-accession migration, but the fact remains that most of them build on qualitative research frameworks and/or are of ethnographic character [Ryan et al., 2009; McGhee, Heath, Trevena, 2013; Cieślińska, 2014; Isański, Mleczko, Seredyńska-Abou Eid, 2014; Main, 2014; Mostowska, 2014]. To the best of our knowledge, there is no research work that investigates Polish emigration patterns with the help of thresh-
old models. Thus, the paper proposes a new approach to (or a new tool for) analyzing the mechanisms that underlie migratory processes, thereby advancing our understanding of the dynamic of cross-border labour mobility and making a contribution to the existing literature of the subject. The structure of the paper is as follows. The next part focuses on the econometric models used in the study. Description of the data and discussion of the results ensue. In this part, emphasis is placed on (the implications of) the threshold values of the transition variables. The paper concludes by summarizing the argument and suggesting future research directions.

1. Overview of the econometric models used in the study

The basic TAR model used in this study can be presented in the following way:

\[
y_t = \sum_{j=1}^{r} (\beta_j + h_j \epsilon_t) \mathbb{I} \{ c_{j-1} < z_{t-d} \leq c_j \} = \\
\begin{cases} 
\beta_{10} + \beta_{11} y_{t-1} + \ldots + \beta_{1p} y_{t-p} + h_1 \epsilon_t & \text{for } z_{t-d} \leq c_1 \\
\beta_{20} + \beta_{21} y_{t-1} + \ldots + \beta_{2p} y_{t-p} + h_2 \epsilon_t & \text{for } c_1 < z_{t-d} \leq c_2 \\
\ldots \\
\beta_{r0} + \beta_{r1} y_{t-1} + \ldots + \beta_{rp} y_{t-p} + h_r \epsilon_t & \text{for } c_{r-1} < z_{t-d}
\end{cases}
\]

where: \(X_{yp} = [y_{t-1} \ldots y_{t-p}]\) is the vector of the endogenous variables, \(\beta_{jp} = [\beta_{j0} \beta_{j1} \ldots \beta_{jp}]\) are the estimated state-dependent parameters, \(p_j\) is the order of \(j\)-state autoregression, \(j = 1, \ldots, r\) is the index of states, \(\mathbb{I}\{\cdot\}\) is the Heaviside threshold function, \(-\infty = c_0 < c_1 < \ldots < c_{r-1} < c_r = \infty\) are the threshold parameters, \(\epsilon_t \sim i.i.d.\) (0.1) is an identically independently distributed random variable, \(z_{t-d}\) is a threshold variable, and \(d\) is the parameter of delay of transition.

Such a model estimates the \(\beta_{jp}\) matrix as well as the number of states \(r\), the orders of autoregression of particular states \(p_j\), threshold parameters \(c\) and the parameters of delay \(d\) of the threshold variable. In this study, we used six different variants of Model (1). The first variant, rooted in the hypothesis of emigration hysteresis, assumes that the equilibrium level (or rate dependent on the degree of integration) of emigration depends on its level in the past. In other words, we assume that the delayed increase in emigration levels is the endogenous variable which initiates a transition between states \(z_{t-d}\). Thus, this model can be classified as the SETAR or M TAR model. The other five variants use the following exogenous variables: (a) purchasing power parity (PLN/EUR), (b) nominal PLN/EUR exchange rate, (c) domestic unemployment rate, (d) GDP growth, and (e) average wage.
While using SETAR models, the discrete data generating (DGP) process is assumed. In practice, however, such variables as unemployment or wages are of autoregressive character, which suggests smooth (rather than sudden) changes. For that reason, we will additionally estimate STAR models with the first order log function as a transition function. These are LSTAR models and can be presented in the following way:

\[ y_t = X_t \beta_1 + X_t \beta_2 \cdot F(S_t; \gamma, c) + \epsilon_t \]  

where \( X_t, \beta_1, \beta_2, \epsilon_t \) are the same as in Model (1), \( F(S_t; \gamma, c) \) is a transition function, and \( S_t \) is a threshold variable. In the present study it is assumed that the transition function is the first order log function: \( F(S_t; \gamma, c) = 1 - e^{-(S_t - c)^2}, \gamma > 0 \), threshold variables \( S_t \) are the same variables used in SETAR models.

2. Description of the data

Our annual dataset, which was derived from the Central Statistical Office of Poland, Eurostat and the National Bank of Poland, comes from the period of 2002–2011. We know that our sample is not long, but due to the political and systemic transition started in 1989, a longer dataset is not available. All the data, except those for emigration and purchasing power parity (PPP), were available in monthly frequency. We used Ecotrim 1.01. to disaggregate emigration data into monthly frequency. Crucially, to disaggregate PPP data, we used our own procedure based on econometric modelling with time-varying parameters. As a result, we obtained a 2002m1–20011m12 dataset.

The PPP model with time-varying parameters can be presented in the following way:

\[ e_t = \alpha + \beta_{1t} p_i + \beta_{2t} p_i^* + \xi_t; \quad \xi_t \sim N(0, \sigma_{\xi}^2); t = 1, \ldots, T \]  

where \( p, p^* \) are logs of domestic and foreign price indices, \( e \) log of the nominal PLN/EUR exchange rate (small letters denote logs of a particular variable). Referring to the relative and strong PPP hypothesis [MacDonald, 1999], we imposed in model (3) the following restriction \( \beta_2 = -1 \); hence, the PPP model (in steady state representation) can be formulated as follows:

\[ e_t + p_t^* - \alpha = \beta_{1t} p_t + \xi_t \]  
\[ \beta_{1t} = \beta_{1t} + \zeta_t \]  

We estimated Model (4)-(5) using the Kalman filter (our procedure in R code is available on request), calculated theoretical exchange-rate values and tested the stationarity of residual values of Model (3). Unit root and stationarity statistics tests were as follows: ADF(-4.977/[0.001]) and KPSS(0.1109). The results suggest
that Relation (3) can be regarded as CI(1.1), which allows us to use the model with time-varying parameters to interpolate annual PPP data to monthly frequency (for details, see: [Zientara, 2012]).

3. Discussion of the results

We provided ADF and KPSS tests for the unit roots and stationarity of our time series. We used EViews 6.0. and the Schwarz Bayes criterion (BSC) to determine autoregressive structures.

Table 1. Results of unit root (ADF) and stationarity (KPSS) tests

<table>
<thead>
<tr>
<th>Variable</th>
<th>ADF/statistics/[prob]</th>
<th>KPSS/statistics</th>
<th>Integration</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>level first difference</td>
<td>level first differenc</td>
<td></td>
</tr>
<tr>
<td>emigration</td>
<td>-2.3279/[0.0974]</td>
<td>-5.3357/[0.0000]</td>
<td>2.5571 0.2633</td>
</tr>
<tr>
<td>GDP</td>
<td>-0.0435/[0.9613]</td>
<td>-6.081/[0.0000]</td>
<td>0.8422 0.0664</td>
</tr>
<tr>
<td>PPS</td>
<td>-2.3788/[0.1380]</td>
<td>-10.215/[0.0000]</td>
<td>1.2267 0.0891</td>
</tr>
<tr>
<td>unemployment</td>
<td>-0.6944/[0.8736]</td>
<td>-4.9876/[0.0001]</td>
<td>2.3872 0.1541</td>
</tr>
<tr>
<td>wages</td>
<td>0.4487/[0.8631]</td>
<td>-4.8851/[0.0001]</td>
<td>2.4501 0.1378</td>
</tr>
<tr>
<td>PLN/EUR</td>
<td>-1.5338/[0.3753]</td>
<td>-7.2307/[0.0000]</td>
<td>0.7588 0.1117</td>
</tr>
</tbody>
</table>

* Critical KPSS test values (intercept): 1% – 0.739; 5% – 0.463; 10% – 0.347; KPSS (intercept + trend): 1% – 0.216; 5% – 0.146; 10% – 0.119.

Source: Own elaboration using Eviews 6.0.

Given the fact that all the variables are I(1), ARIMA and TAR models were estimated for first differences of particular variables. So, these models show the pace of emigration rather than its level. Then, we tested the linearity of emigration models, using BDS [Brock et al., 1986] and Terasvirta [1994]. The results are presented in Table 2.

Table 2. Results of Terasvirta ($\chi^2$), Terasvirta (F) and BDS tests

<table>
<thead>
<tr>
<th>Test</th>
<th>Statistic</th>
<th>prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>Terasvirta ($\chi^2$)</td>
<td>81.3427</td>
<td>3.237•10^{-10}</td>
</tr>
<tr>
<td>Terasvirta (F)</td>
<td>16.3671</td>
<td>3.556•10^{-8}</td>
</tr>
<tr>
<td>BDS</td>
<td>69.3490</td>
<td>2.261•10^{-14}</td>
</tr>
</tbody>
</table>

Source: Own elaboration using R-Crane.

Results suggest H0 rejection on every commonly used significance level, which means the non-linear character of the relationship of $\Delta Em_t$. This justifies the
use of SETAR and LSTAR models. Our next step was to compare the *ex post* prognostic properties of the ‘best’ ARIMA model with the properties of the above-mentioned non-linear models (the ‘best’ model meets the BSC out of all the models with no autocorrelations in residuals). The ‘best’ ARIMA model turned out to be the following ARIMA(3,1,0):

\[
\Delta E_{mt} = 1.1489 + 2.8076 \Delta E_{m,t-1} - 2.6309 \Delta E_{m,t-2} + 0.8226 \Delta E_{m,t-3} + \epsilon_t \tag{6}
\]

where in parentheses asymptotical t-Student values are reported. Except for the intercept, all parameters are significant. All the modules of the roots of the characteristic equation lie out of the unit circle (the modules are 1.1307; 1.0369; 1.0369). Thus we can regard \(\Delta E_{mt}\) as a stationary variable, which means that we can forecast it by means of Model (6). Subsequently, we estimated TAR models, which allow for multiple equilibriums. Given the fact that we estimated six different versions of the TAR model, the final results are not presented individually, but in Table 3. It has to be noted that in each of the six models we found two equilibriums (or \(r = 2\)).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Autoregression order in lower equilibrium ((p_1))</th>
<th>Autoregression order in upper equilibrium ((p_2))</th>
<th>Delay of transition variable ((d))</th>
<th>Threshold value (z_{rd})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emigration</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>10,570[individuals]</td>
</tr>
<tr>
<td>GDP</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3.57[% annual]</td>
</tr>
<tr>
<td>Unemployment</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>-0.064[% points]</td>
</tr>
<tr>
<td>PPP</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>0.00361[PLN]</td>
</tr>
<tr>
<td>Wages</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>10.42[PLN]</td>
</tr>
<tr>
<td>PLN/EUR</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0.00551[PLN]</td>
</tr>
</tbody>
</table>

Source: Own elaboration using R-Crane.

While interpreting the results, one should highlight the following aspects:

- in each TAR model (both in the endogenous and exogenous transition function) we found two equilibriums. This facilitates interpretation: models indicate two equilibriums dependent on the threshold value of the transition function (or the transition variable). Since the number of equilibriums is a priori unknown and is estimated within the model, the identical \(r = 2\) results obtained for different transition variables suggest their correctness;

- delay in the adjustment of emigration pace is fastest (= one month) when wages are used as a transition variable (see: Table 3). It follows that the pace of emigration of Poles reacts fastest – or *switches to a higher level* – to an increase in wages (that is, to the variable itself rather than to the very phenomenon). In practice, this means that if a monthly nominal growth of wages is lower than...
PLN 10.42, the pace of emigration switches to a higher level (it is important to note that the threshold value – 10.42 – is lower than an average monthly wage growth, which is PLN 11.31 in the 2002-2011 period);
– delay in the adjustment of emigration pace is slower (= two months) when GDP, unemployment, wages and exchange rate are used as threshold values (Table 3). This means that if a monthly GDP growth is lower than 3.57% annually or a monthly increase in unemployment is higher than -0.064 p.p. or PPP is higher than PLN 0.00361 or the PLN/EUR exchange rate is higher than PLN 0.00551, then the pace of emigration switches to a higher level (note that average values of the above transition variables are, respectively, 4.23%, -0.084 p.p., PLN 0.00374 and PLN 0.00543).
– delay in the adjustment of emigration pace is slowest (= three months) when endogenous emigration pace is used as a transition variable. This means that if a monthly growth in emigration is higher than 10,570 individuals, then the pace of emigration switches to a higher level (note that the average monthly value of this transition variable is 11,410 individuals). It is worrying, from a certain point of view, that even though the threshold value is lower than the monthly average, the switch to a higher level is triggered. What is more, we interpret the three-month delay in the following way: when a new emigrant leaves Poland, this – after 3 months – causes other people (usually, relatives and/or acquaintances) to emigrate, too.

The next step of our inquiry was to compare the prognostic (forecast) accuracy of the ARIMA model with that of the TAR models. To that end, we divided our sample into two sub-samples: 2002m1-2011(m12-h) and 2011(m12-h+1)-2011m12 (h is a forecast horizon, h = 12, h = 24). We estimated ARIMA and TAR models, using the first sub-sample, and we formulated the forecast for the second sub-sample. Then we verified the ex post forecast accuracy by comparing ME and RMSE errors and by conducting the Diebold-Mariano test (see: Tables 4 and 5).

Table 4. Comparison of ME and RMSE errors

<table>
<thead>
<tr>
<th>Model/Threshold variable</th>
<th>h=12</th>
<th>h=24</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ME</td>
<td>RMSE</td>
</tr>
<tr>
<td>ARIMA</td>
<td>-1.506</td>
<td>2.125</td>
</tr>
<tr>
<td>emigration</td>
<td>-0.890</td>
<td>1.455</td>
</tr>
<tr>
<td>GDP</td>
<td>0.742</td>
<td>1.121</td>
</tr>
<tr>
<td>unemployment</td>
<td>0.504</td>
<td>0.604</td>
</tr>
<tr>
<td>PPP</td>
<td>-0.249</td>
<td>0.363</td>
</tr>
<tr>
<td>wages</td>
<td>-0.529</td>
<td>0.706</td>
</tr>
<tr>
<td>PLN/EUR</td>
<td>-1.098</td>
<td>1.408</td>
</tr>
</tbody>
</table>

Source: Own elaboration using R-Crane.
Table 5. Results of the Diebold-Mariano test

<table>
<thead>
<tr>
<th>Threshold variable</th>
<th>h=12</th>
<th>h=24</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>stat/prob</td>
<td>stat/prob</td>
</tr>
<tr>
<td>emigration</td>
<td>3.221/0.000</td>
<td>3.552/0.000</td>
</tr>
<tr>
<td>GDP</td>
<td>2.337/0.004</td>
<td>3.244/0.000</td>
</tr>
<tr>
<td>unemployment</td>
<td>2.553/0.005</td>
<td>2.988/0.001</td>
</tr>
<tr>
<td>PPP</td>
<td>2.547/0.006</td>
<td>2.994/0.000</td>
</tr>
<tr>
<td>wages</td>
<td>2.110/0.009</td>
<td>2.527/0.008</td>
</tr>
<tr>
<td>PLN/EUR</td>
<td>3.158/0.000</td>
<td>3.681/0.000</td>
</tr>
</tbody>
</table>

Source: Own elaboration using R-Crane.

As regards the forecast horizon $h=12$, the TAR model with PPP as a threshold variable (the ‘best’ TAR model) provides forecasts for which mean forecast error (ME) is only 16.5% of the ARIMA value and for which root mean square error (RMSE) is 17.1% of the ARIMA value (see: Table 4). The results reported in Table 4 and Table 5 show that TAR models provide more accurate forecasts than ARIMA models, regardless of the forecast horizon. This seems to confirm the results of the non-linearity hypothesis and the multiple equilibriums hypothesis.

Conclusions

Polish post-accession migration, which has taken on considerable proportions, constitutes fertile ground for exploring various aspects of cross-border labour mobility. This study has confirmed our *a priori* assumption that SETAR models are more adequate than ARMA/ARIMA models for estimating the pace and rate of Polish emigration. At the same time, we have estimated the threshold values of selected transition variables. Of special interest is the finding that – when wages are used as a transition variable – delay in the adjustment of emigration pace is fastest (one month). On the other hand, our results also add substance to the claim that prior emigration is an important pull factor. Therefore, our research also carries practical significance: knowledge of the threshold values as well as of the ‘influence’ of concrete transition variables can help make more informed policy choices.

Relatedly, our method could be applied to data from other economies – both within the EU and outside it – which have also seen massive outflows of people. In particular, this goes for countries located in South and South-East Asia [Oziewicz, 2007]. Thus, one line of future inquiry might focus on emigration patterns in such countries as Burma, Bangladesh or the Philippines (which continue to experience...
large outflows of individuals who seek employment in more advanced economies in the region: Indonesia, Malaysia, Thailand or in the Gulf countries. In the first half of 2015, the surge of migrants in South-East Asia reached a level that was critical and alarming [The Guardian, 2015]). It has become evident that the problem is wider and could be named global. This could expand the body of research into the dynamics of migratory processes and, simultaneously, provide insights of practical character. Admittedly, it needs to be stressed that the principle of free movement of labour is a prerequisite of applicability of our approach. And, as is widely known, this is not the case across South-East Asia. Besides, given considerable cultural differences between Europe and Asia, the question arises of whether other contextual (cultural) factors might play a more important role than purely economic ones (as is assumed in this study). Future researchers might also wish to explore the implications of this problem.

References

Brock W.A., Dechert W.D., Scheinkman J., 1986, A Test for Independence Based on the Correlation Dimension, unpublished working paper, Department of Economics, University of Wisconsin, University of Houston and University of Chicago.


McGhee D., Heath S., Trevena P., 2013, Post-Accession Polish Migrants: Their Experiences of Living in Low-Demand Social Housing Areas in Glasgow, Environment and Planning A, no. 2.


