PALKANSAAJIEN TUTKIMUSLAITOS
LÖNTAGARNAS FORSKNINGSINSTITUT
LABOUR INSTITUTE FOR ECONOMIC RESEARCH

TYÖPAPEREITA | WORKING PAPERS 330

Identification of fiscal SVARs in small open economies using trading partner forecast errors as instruments*

Henri Keränen**
Sakari Lähdemäki***

* We thank Ilkka Kiema for his support throughout this project and Roni Mäkelä for his help in gathering our dataset of past forecast errors. We would also like to thank Samu Kärkkäinen for his useful comments. We gratefully acknowledge the funding provided by Palkansaajasäätiö.

PALKANSAAJA-SÄÄTIÖ

** henri.keranen@labour.fi, Labour Institute for Economic Research
*** sakari.lahdemaki@labour.fi, Labour Institute for Economic Research

Abstract

We identify structural vector autoregressions with external instruments (SVAR-IV) to study the dynamic effects of fiscal policy. Our main contribution is a novel instrument for aggregate output shocks of a small open economy. Unexpected shocks in domestic output are proxied by forecast errors of professional forecasters in trading partner economies. Our instrument relies on two key assumptions. Firstly, unexpected changes in trading partners are correlated with unexpected shocks of an open economy (relevance). Secondly, unexpected fiscal shocks of a small economy are unrelated with the forecast errors of its trading partners (exogeneity). Test results show that this instrument is relevant. We find suggestive evidence that our instrument is more credibly exogenous than the prevailing TFP instrument. We apply our instrument to estimating fiscal SVAR models of two countries, Canada and Finland, and find that estimates of the spending multiplier are sensitive to conventional identification assumptions.

Keywords: Fiscal policy, Fiscal multiplier, SVAR-IV, Small open economy
JEL Codes: E62, C26, C32

Tiivistelmä

Tutkimuksessa tarkastellaan finanssipolitiikan dynaamisia vaikutuksia SVAR-mallilla, jossa rakenteelliset parametrit identifioidaan käyttämällä ulkoisia instrumentteja. Tutkimuksen tärkein kontribuutio on uusi instrumentti tuotantoshokeille. Instrumentti muodostetaan hyödyntäen ammattimaisten ennustajien ennustevirheitä kauppakumppanimaissa. Instrumentin soveltuvuus perustuu kahteen keskeiseen oletukseen. Ensinnäkin tärkeiden vientimaiden tuotannon odottamattomat muutokset korreloivat pienen avoimen talouden tuotantoshokkien kanssa (relevanttius). Toiseksi pienen avoimen talouden fiskaaliset shokit eivät ole yhteydessä sen vientimaiden odottamattomiin tuotannon muutoksiin (eksogeenisuus). Testitulokset osoittavat, että ehdotettu instrumentti on relevantti. Tulokset tarjoavat myös viitteellisiä todisteita siitä, että eksogeenisuusoletus saattaa olla uskottavampi esitetylle kauppakumppani-instrumentille kuin TFP-instrumentille, jota on aiemmin käytetty kirjallisuudessa. Ehdotettua instrumenttia sovelletaan kahden eri maan, Kanadan ja Suomen, fiskaalisia SVAR-malleja estimoitaessa. Instrumenttia käytettäessä havaitaan, että estimoidut finanssipoliittiset kertoimet ovat erityisen herkkiä fiskaalisten SVAR-mallien perinteisiä identifiointirajoituksia koskeville oletuksille.

Avainsanat: finanssipolitiikka, finanssipoliittinen kerroin, SVAR-IV, pieni avotalous

Palkansaajien tutkimuslaitos
Labour Institute for Economic Research
Arkadiankatu 7
00100 Helsinki

Työpapereita | Working Papers 330
ISBN 978-952-209-189-5(pdf)
ISSN 1795-1801 (pdf)

www.labour.fi

Helsinki 2020

1 Introduction

After the financial crisis of 2008 there has been a resurgence of interest in the role of fiscal policy as an essential tool of macroeconomic stabilization (see Blanchard and Summers (2017) and Ramey (2019)). This resurgence coincides with a period of lackluster growth, negative output gaps and inflation below target levels which have characterized the decade after the crisis. All this even under the unconventionally aggressive monetary policy that has driven policy rates near or below zero and massively increased balance sheets of central banks. At the same time it has been acknowledged that the low interest rate environment has arguably created more room for active fiscal policy (see Blanchard, 2019).

There is an enduring discussion on whether or not government spending is an effective way to stimulate an economy. As articulated in Nakamura and Steinson (2018) there are two main sources of aggregate evidence on fiscal multipliers: evidence from vector autoregressions and wars (narratively identified exogenous shocks based on US military spending). Over the recent decade, the SVAR (structural vector autoregression) literature has seen a development into a direction where the identification of the structural shocks is achieved with the use of external instruments in SVAR-IV (or Proxy-SVAR) models (Stock and Watson, 2018). In this paper we follow this approach in the identification of fiscal policy shocks. More specifically, we follow the strategy of using non-fiscal instruments in identifying fiscal shocks in a macro-fiscal SVAR-IV-model as Caldara and Kamps (2017) propose.[1]

Our main contribution is a novel instrument for output shocks that can be used to identify the contemporaneous output elasticities of endogenous fiscal variables. These elasticites are central in determining the effects of fiscal stimulus, for example, in the SVAR-model of Blanchard and Perotti (2002). We propose to use professional forecast errors of trading partner economies as a proxy for output shocks. These errors are arguably unrelated to government spending and tax shocks of a small open economy. More specifically, our instrument relies on the following assumptions. Firstly, unexpected changes in trading partners are correlated with unexpected shocks of an *open* economy (relevance). Secondly, unexpected fiscal shocks of a *small* economy are unrelated with the forecast errors of its trading partners (exogeneity). Arguably, the more open the economy is and the smaller the

[1] Following the SVAR model of Blanchard and Perotti (2002) much of the macro-fiscal research has relied on using assumptions about timing restrictions in the real-world setting of fiscal policy to identify fiscal shocks. Another prominent identification strategy has been to use sign restrictions as in Mountford and Uhlig (2009). Furthermore, another branch of the literature has utilized narratively identified plausibly exogenous policy changes or other shocks in the identification of the effects of fiscal policy. See, for example, Romer and Romer (2010) and Mertens and Ravn (2013) in the context of tax policy changes and Ramey (2011) and Ramey and Zubairy (2018) in which a military-news variable is constructed to identify plausibly exogenous variation in government spending. See Ramey (2019) for a more comprehensive summary of the macro-fiscal literature.

economy is, the more confident one can be of these assumptions to hold.

More specifically, we need not be concerned whether some variation in the instrument is explained by common shocks. This is acceptable as long as this common variation in the domestic economy and its trading partners is not due to fiscal policy shocks of the small domestic economy. Implicitly we are thus assuming that there is no spillover from the unexpected fiscal shocks of a small economy to the forecast errors of other countries. We argue that this is a reasonable assumption. Consider, for example, Finland where the pass-through from a Finnish fiscal policy shock to other economies would have to implausibly large for it to matter for the forecast errors of its often much larger trading partners.

Preceding SVAR-IV literature has considered the utilization adjusted total factor productivity (TFP) series of Fernald (2014), which for example Caldara and Kamps (2017) utilize, as the instrument of choice for output shocks. Compared with this prevailing TFP based instrument, the instrument we propose is formed straight from observable data and does not require any strong structural assumptions.[2] Furthermore, the analysis in Angelini et al. (2020) raises some concerns that the utilization adjusted TFP series of Fernald (2014), if used as an instrument, does not necessarily fulfill the exogeneity assumption. Based on a proxy for fiscal policy shocks we provide suggestive evidence that our instrument seems to fulfil the exogeneity assumption more reliably. Moreover, it has been shown by Kurmann and Sims (forthcoming) that revisions made to the utilization adjusted TFP series of Fernald (2014) have altered it remarkably.[3] Given these concerns, we see that the instrument we propose is a viable alternative to the TFP instrument in a world of scarce instruments.

We apply the new instrument to two different types of small open economies: Canada and Finland. While the intuition behind the instruments for Canada and Finland is the same, the construction of the instrument for these two differs. Roughly speaking, more than 70 percent of Canada's exports are destined to the USA. Therefore, the GDP forecast errors for the US economy can be expected to be a good predictor for unexpected movements in Canada's GDP. Contrary to Canada, Finland's exports are more diversified among a number of destination countries. The top 10 export countries form roughly 60 percent of Finland's total exports, whereas at least top 15 countries are needed to form more than a 70 percent share of total exports.[4] To form an instrument for Finland that covers roughly a similar share of exports as the instrument for Canada, we gather GDP and import forecast errors

[2]Note also that our instrument is not a narrative instrument as, for example, in Romer and Romer (2010), Mertens and Ravn (2013) or Ramey and Zubairy (2018). That is, we do not choose what shocks are exogenous and what are not. We only assume that the forecast we use are reasonable and that the resulting forecast errors can proxy unexpected shocks at the quarterly frequency.

[3]Note that the utilization adjusted TFP series is readily available only for USA. See Comin et al. (2020) for utilization adjusted TFP series for major European countries.

[4]These figures are based on averages of export shares over the years 2013-2015.

for several countries that are important trading partners for Finland and weight these errors by their respective export shares. That is, for Finland our instrument is a weighted average of the trading partners forecast errors.

Furthermore, Canada and Finland differ in the way they fit the definition of a small open economy over the sample we study them. Canada is clearly an open economy while it is perhaps small only when compared to its neighboring USA. However, traditionally Canada is considered to be a small open economy.[5] Finland on the other hand is economically clearly small. Yet, Finland's trade has had some unique features before the collapse of the Soviet union.[6] Anyway, it is not crucial that the studied economy meets perfectly the definition of a small open economy in the traditional sense. Our claim is that the characteristics of a small open economy best facilitate the validity of our proposed instrument.

With our instrument, we identify fiscal SVARs for both Canada and Finland in a simple 3-variable framework similar to Blanchard and Perotti (2002) and Caldara and Kamps (2017). While estimating the SVAR models, we also test for the relevance of the instruments. Namely, we test the possible weakness of the instruments with the efficient F-test proposed by Montiel Olea and Pflueger (2013). The instruments in our main specifications are not weak in the sense of the rule of thumb that the efficient F statistic > 10 (see Andrews, Stock and Sun, 2019). We thus show that the test results support the validity of our instrument for both Finland and Canada.

We also examine the resulting impulse responses and cumulative fiscal multipliers across four different SVAR-IV identification schemes we consider. We find that the impulse responses related to government spending are sensitive to the zero restriction of the output elasticity of government spending inherent in the Blanchard and Perotti (2002) identification that can be relaxed with the use of our instrument similarly as in Caldara and Kamps (2017). As a result of this sensitivity, also the estimates of the government spending multiplier vary depending on the identification scheme used. Anyway, we find that the 2SLS estimates of output elasticities of fiscal variables are quite robust to a battery of changes in the reduced form VAR specification.

The rest of the paper is structured as follows. In section 2 we present how we gather and construct our data. Section 3 describes our empirical strategy relating to the specification of the SVAR-IV-model. In section 4 we examine the existing TFP instrument, discuss the

[5]For example, the Bank of Canada uses models based on the small open economy assumption in their projections, see Dorich et al. (2013). See also Awokuse (2003).

[6]This is since the share of Finland's trade (roughly a 20 % fraction) that was directed towards Soviet Union was not based on free markets and the share of trade (roughly a 80 % fraction) that was directed towards the western countries was free market oriented. For more, see, for example, Gorodnichenko, Mendoza and Tesar (2012) and Gulan, Haavio and Kilponen (2019).

validity of our proposed instrument and provide test results relating to this matter. In section 5 we apply our instrument to estimating fiscal SVAR-IV-models for both Canada and Finland and discuss the resulting impulse responses. We also comment on the government spending multiplier estimates derived from the estimated SVARs. Finally, section 6 concludes.

2 Data

Endogenous variables.— The data for our VAR endogenous macro variables is from the statistical agencies of Finland and Canada.[7] For Canada our sample period is 1961Q1-2018Q4. To obtain a comparatively long time series for Finland, we combine Statistics Finland data which is available from 1999Q1 onwards with the dataset used in Virkola (2014) that contains quarterly revenue statistics as well as the other considered variables from 1975Q1 onwards. This data is originally also from Statistics Finland. The sample period for Finland is 1975Q1-2018Q4. For both countries, government spending and investment are deflated by their respective deflators while the revenue side is deflated by the GDP deflator as is output. All data are seasonally adjusted by the statistical agencies. The upper panels of Figure 1 plot our sample of these endogenous variables for Canada and Finland respectively.

Instrument data.— For our instrument we collect data on past macroeconomic forecasts of professional forecasters. The forecasts we consider are often reported in levels. Since the level forecasts of real variables rely on base years that change over forecast vintages and are also conditional on the then available and later revised information on past levels of the variables, we transform all level forecasts to log-difference forecasts as follows. Let $F_t[.]$ denote a forecast operator and v_{t+1} is the value of variable v in period $t+1$. Then $F_t[v_{t+1}]$ is the forecast of the value of v in period $t+1$ made in period t. We use the forecasted log-difference

$$F_t[ln(v_{t+1}) - ln(v_t)]$$

to represent the forecast made at time t. Forecast errors of variable v at time $t+1$ are then obtained as the difference between the realized log-difference of v_{t+1} and the forecasted one, i.e.

$$\{ln(v_{t+1}) - ln(v_t)\} - \{F_t[ln(v_{t+1}) - ln(v_t)]\}.$$

For Canada, we use quarterly forecasts of the US economy from the Survey of Professional forecasters (SPF).[8] In SPF, each quarter after the first release of the previous quarter's GDP

[7] These agencies are Statistics Finland (quarterly national accounts and general government revenue and expenditure by quarter) and Statistics Canada (CANSIM database) respectively.

[8] All forecast vintages of the Survey of Professional forecasters are readily available from the Philadelphia Fed website `https://www.philadelphiafed.org/research-and-data/real-time-center/`

figure, a panel of professional forecasters are asked to provide forecasts of several macro-variables of the US economy. The forecast is released in the middle of each quarter. In constructing the instrument, we use the mean forecast of one quarter ahead real GDP. The real GDP forecast is reported in levels and thus we transform it to a log-difference forecast as outlined above. The time series of the resulting instrument for Canada is plotted in lower panel of Figure 1.

For Finland, our data for the instrument is from the OECD Economic Outlooks (EOs) which contain both the then available history and contemporaneous forecasts of several macroeconomic variables for a number of different countries. OECD EOs are published twice a year in June and December. For EOs published before 2003S2, only annual and semi-annual forecasts are available whereas from 2003S2 onwards the EOs contain both annual and quarterly forecasts for a large subset of variables. When quarterly forecasts are not available, we interpolate the semiannual growth rate (log-difference) forecast over the two quarters. In other words, we assume that the semiannual growth rate forecast is constant across the period, so that we can simply divide the semi-annual log-difference by two into two quarterly log-differences.

Since, unlike for Canada, no single country has an overwhelming share in Finnish exports, we combine forecast errors from a number of countries into a single instrument. We also weight trading partner forecast errors by their share in Finnish exports. The export weights data is from OECD International trade statistics that contain quarterly data. Since past vintages of OECD EOs have a better coverage for trade forecasts than for GDP forecasts for a number of countries, we consider both the forecast errors of GDP growth as well as of import growth in building our instrument for Finland. The lower panel in Figure 1 plots the resulting time series for the instrument using all available countries in the sample. For comparison, we construct both import based and GDP based instruments also for Canada using the data in OECD EOs analogously to Finland. Nevertheless, the SPF based GDP instrument is considered as the prime instrument for Canada.

[Figure 1 here.]

3 Empirical strategy

Reduced form VAR.— Our baseline VAR is a model of a vector $y' = [g, r, x]$ of three endogenous variables of interest: general government consumption and investment (g) and

`survey-of-professional-forecasters/.`

government net revenue (r) and GDP (x).[9] All variables enter the model in real per capita terms and in natural logarithms. Following Caldara and Kamps (2017), we also linearly detrend each endogenous variables prior to estimating the VAR model in our baseline specification.[10] Accordingly, we specify our reduced form VAR model as follows:

$$y_t = c + \sum_{i=1}^{p} A_i y_{t-i} + u_t,$$

where p is lag length, $c \in \mathbf{R}^{3\times 1}$ is a vector of constants, $A_i \in \mathbf{R}^{3\times 3}$ are the autoregressive coefficient matrices and $u_t \in \mathbf{R}^{3\times 1}$ are the model residuals. We estimate the reduced form VARs with equation-by-equation OLS and conduct the related impulse response analysis as is customary.

In lag length selection we rely on both the standard information criteria as well as on the partial autocorrelation function of the residuals in order to specify a model that both contains enough information and has no autocorrelation in the residuals. For both Finland and Canada, our baseline specification has a lag length of 5.[11] We consider the robustness of our results using different lag lengths in the reduced form VAR in later sections.

The confidence intervals we show are based on residual-based moving block bootstrap proposed by Brüggemann, Jentsch and Trenkler (2016). See appendix A for discussion and details. The fiscal multipliers we derive are the so called present value cumulative multipliers similar to what Mountford and Uhlig (2009) use. See appendix G for discussion and details.

Structural form.— . We next describe in detail how we identify the structural models. Across the paper we consider four different types of structural specifications, all of which are close relatives to each other but differ in certain respects. Underlying all of these structural forms is a Blanchard and Perotti (2002) type identification scheme which is a combination of assumptions in both the elasticities between contemporaneous values of endogenous variables and in the direct responses of endogenous variables to the structural shocks.

[9] Note that g differs from general government expenditures and r from general government revenues. This is since transfer payments, that are an expenditure, are included in the revenue side as a negative factor.

[10] Applied VAR papers vary in how they approach trends in the VAR. In Blanchard and Perotti (2002) a linear and a quadratic trend are included in the main specification. Caldara and Kamps (2017) detrend some of their endogenous variables. Mertens and Ravn (2013) do not add deterministic trends in their main specification, while they test the robustness of their results with such a specification. Gertler and Karadi (2015) do not add deterministic trends in their main specification. According to Kilian and Lüthkepohl (2017), while a VAR in levels is asymptotically valid even under true cointegration relations, its finite sample bias can be considerable. This bias is even more severe when a deterministic trend is included in the model.

[11] Typically much of the applied VAR literature uses four lags with quarterly datasets. However, in our case with 4 lags there is some autocorrelation left in the residuals for both countries. To ensure there is no autocorrelation in the residuals we include 5 lags in our baseline specification of the reduced form VAR. In appendix F we also examine whether the instrument is robust to different choices relating to the specification of the reduced form VAR.

Consider the following system of equations:

$$A_0 u_t = B e_t,$$

where $A_0 \in \mathbf{R}^{3\times3}$ contains the contemporaneous elasticities between endogenous variables, $u_t \in \mathbf{R}^{3\times1}$ are the VAR residuals, $B \in \mathbf{R}^{3\times3}$ is a coefficient matrix and $e_t \in \mathbf{R}^{3\times1}$ are the i.i.d. mean zero structural shocks. The elements of matrix A_0 and B represent the relations between these exogenous structural shocks and VAR residuals. The identification schemes we consider can all be represented in this AB-form:

$$\underbrace{\begin{bmatrix} 1 & 0 & -a_{gx} \\ 0 & 1 & -a_{rx} \\ -a_{xg} & -a_{xr} & 1 \end{bmatrix}}_{A_0} \begin{bmatrix} u^g \\ u^r \\ u^x \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & b_{gr} & 0 \\ b_{rg} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{B} \begin{bmatrix} e^g \\ e^r \\ e^x \end{bmatrix} \tag{1}$$

that nests the typical Blanchard and Perotti (2002) (BP) identification scheme as a special case.

The relationships in the original BP identification are assumed to be of the following type:

$$u^g = a_{gx} u^x + b_{gr} e^r + e^g \tag{2}$$

$$u^r = a_{rx} u^x + b_{rg} e^g + e^r \tag{3}$$

$$u^x = a_{xg} u^g + a_{xr} u^r + e^x \tag{4}$$

The BP identification relies on a priori knowledge and assumptions about the elasticities of government spending and net revenues to changes in output that are directly inserted to the A_0-matrix. Parameter a_{rx} is constructed from estimates of the automatic response of net revenues to output. In addition, it is assumed that within quarter contemporaneous response of the policymaker to unexpected output, a_{gx}, is zero. This is due to the data being quarterly and decision lags being assumed to be longer than that.

In effect, BP identification is done recursively equation-by-equation. Parameters a_{gx} and a_{rx} are given their a priori assumed values and either b_{gr} or b_{rg} in the B-matrix is set to zero. For example, if b_{rg} is set to zero, that is, net revenues are not contemporaneously affected by structural government spending shocks, then structural net revenue shocks are outright uncovered from the reduced form residuals as $e_t^r = u_t^r - a_{rx} u_t^x$. These shocks can then be used to estimate parameter b_{gr} in $u_t^g - a_{gx} u_t^x = b_{gr} e_t^r + e_t^g$ where e_t^g are taken to be the structural government spending shocks. Finally, e_t^g and e_t^r are used as instruments for

u_t^g and u_t^r in equation 4 to obtain structural parameters a_{xg} and a_{xr} via 2SLS.

In contrast to the original BP identification which relies upon external estimates of elasticities that are then applied to the model, we follow Caldara and Kamps (2017) and estimate these contemporaneous elasticities within the model using our proposed instrument. Suppose we have an instrumental variable m_t that satisfies the following conditions:

$$E[m_t e_t^x] = \Gamma \neq 0 \tag{5}$$

$$E[m_t e_t^j] = 0, \ j \in g, r. \tag{6}$$

Given that these conditions hold, we can use m_t as an instrument and estimate the contemporaneous elasticities of the policy variables g and r with respect to output x via 2SLS instead of relying on external estimates of parameters a_{xg} and a_{xr}. Equations (2)-(4) reflect the regressions where from we estimate the structural parameters. However, we follow Stock and Watson (2018) and estimate regressions where instead of the estimated VAR residual we use the endogenous variables and lags of all endogenous variables along side with the derived exogenous series. Using the reduced form residuals instead would lead to the same parameter estimates but one would need to adjust the standard errors ex-post to achieve reliable inference.

As already mentioned, we consider four slightly different identification schemes altogether. Two of these, which we label as BP_g and BP_r, are replications of the ordinary Blanchard and Perotti (2002) identification with the only exception being that we estimate the output elasticity of net revenues using the VAR specification and our proposed instrument instead of assigning the value for this parameter outside the model. The difference between BP_g and BP_r lies in the assumption regarding the ordering of g and r in the VAR as in the original paper of Blanchard and Perotti (2002).

The two other identification schemes are labeled as CK_g and CK_r. The only difference between these and the related BP identifications is that we do not impose a zero restriction on the contemporaneous output elasticity of government spending and investment. The reason for this is that given that we have a valid instrument and a correctly specified VAR-model at hand, we should be able to estimate this parameter in the same manner as we do the parameter for output elasticity of net revenues. Thus, following Caldara and Kamps (2017), we allow these parameters to deviate from zero and estimate them. We summarise the differences between the four different identification schemes in Table 1 and additionally present the four identification schemes in the typical AB-form in Appendix D.[12]

[12] Note that to add comparability between these two schemes (BP and CK) we use the identified structural shock series as instruments in the final regression (equation (4)) following BP in all four identification schemes, rather than use them directly as regressors as is done in CK.

[Table 1 here.]

4 On the validity of the instruments

4.1 Prevailing TFP instrument

Before we proceed to a more detailed discussion of our proposed instrument it is useful to briefly highlight key differences between our proposed instrument and the currently leading instrument for output in the related literature: the quarterly utilization adjusted TFP series of Fernald (2014) which is available for the United States. This series is used as an instrument, for example, in Caldara and Kamps (2017) and in Angelini et al. (2020) when estimating the output elasticity of fiscal variables in a similar setting to ours. In contrast to the TFP instrument, a major advantage of our instrument is that it is derived rather straightforwardly from observable data. This means that the decisions made by the researcher have presumably a lot smaller impact on the final instrument series and thereby also the elasticity estimate.

Construction of the utilization adjusted TFP series of Fernald (2014) involves a number of assumptions. Since TFP growth is that part of aggregate output growth that cannot be explained by changes in inputs, one first needs to at least implicitly decide over a specification of the aggregate production function. Ever since Solow (1957) it has been standard practice to calculate TFP growth as the difference between the growth in aggregate output and a weighted average of input growth rates where each input is weighted by its corresponding factor share. The Solow residual can be calculated this way under the assumption that firms minimize costs and, crucially, make no profits in which case factor shares are the correct weights. In a recent paper, Comin et al. (2020) illustrate how the zero-profit assumption as well as ignoring adjustment costs can bias the calculation of TFP. By relaxing the zero-profit assumption they show that one underestimates long-run TFP growth and overestimates its volatility and cyclicality if one uses the conventional method of calculating TFP growth.

In addition to this, to form the utilization-adjusted TFP series one needs to adjust the TFP series for capacity utilization. This involves additional decisions over how to account for or how to model/proxy the utilization rate. For example, to form their utilization adjusted TFP series Basu, Fernald, and Kimball (2006) as well as Fernald (2014) use changes in hours per worker as their utilization proxy under the assumption that it is one-to-one related to changes in capacity utilization. More recently, Comin et al. (2020) use firm surveys on capacity utilization as a more direct proxy.

Clearly any of the choices made in the process of modeling the utilization adjusted TFP series have an effect on the final TFP series. In fact, as documented in Kurmann and Sims (forthcoming) the revisions made to the utilization adjusted TFP series as a result of new releases of data as well as methodological tweaks made by the author seem to have a remarkable effect on the resulting series of Fernald (2014).

It is also worth noting that when using a TFP series of any kind as an instrument for output one is effectively using a filtered series of output as an instrument for output. This perhaps helps to ensure the relevance of the instrument but also raises concerns about the series' exogeneity. In a fiscal SVAR for instance, for any TFP series to be a valid instrument, the series should be strictly orthogonal to fiscal shocks. Given how the TFP series is constructed, this relies critically on the filtering process to correctly filter out all demand-driven variation in the original output series (under the assumption that fiscal shocks affect demand). Yet, we show in the next subsection that the TFP series of Fernald (2014) can be predicted by government spending forecast errors of professional forecasters. This, we argue, might indicate that the exogeneity assumption does not hold for the TFP instrument. Angelini et al. (2020) allude to a similar issue.

4.2 Proposed trading partner forecast error instrument

The essential assumptions needed for our instrument are the following. Firstly, the professional forecasts are sensible and so the forecast errors capture real unexpected variation in trading partner economies. Secondly, these forecast errors have explanatory power on the unexpected changes of aggregate output of the country of interest. Thirdly, the fiscal policy shocks of this domestic country cannot explain the unexpected variation in its trading partners output. Formally, the latter two assumptions were laid out in equations 5 (relevance) and 6 (exogeneity) respectively.

Relevance.— The argument over why the instrument we propose should satisfy the relevance assumption is simply the following: for a country that is open to trade and whose exports form a large share of its GDP, an unexpectedly good (bad) development in its trading partners has an effect on its output through, for example, growing (declining) the demand for its exports.[13] The intuition of the relevance assumption is quite clear. The remaining question is whether this relation is strong enough to ensure that the problems arising from weak instruments are diluted.

The concept of weak instruments, which has been highlighted lately, is directly related to the relevance assumption. The weak instruments problem springs from the low correlation between the instrument and the corresponding endogenous regressor. The reason for why

[13]Note also that any common/global exogenous output shocks strengthen this relation and does not dilute the exogeneity assumption as long as it is unrelated with the fiscal policy shocks.

the weak instrument problem is highly prominent is that, if the instrument is indeed weak, it hinders inference due to the low efficiency of the estimator. Importantly, it also magnifies the finite sample bias which could mean that the estimates are far off from the true values. Considering the fact that most time series samples are quite finite, extra care should be taken to confirm that the instruments in use are not weak.

To test for the relevance and weakness of an instrument in 2SLS estimation, one can examine the F-statistics of the first stage regression. The test statistic we use is the efficient F-test proposed by Montiel Olea and Pflueger (2013).[14] As argued in Andrews, Stock and Sun (2019), this test should be used instead of the standard F-statistic which assumes homoskedasticity when detecting weak instruments. Accordingly, in the case with only one instrument and one endogenous variable it is sufficient to use the following rule of thumb; efficient F statistic > 10; or to rely on the critical values provided in Stock and Yogo (2005).

The above test and the related rule of thumb can be used to detect weak instruments. This has, as we understand, more to do with the possible bias of the point estimator (location) than inference of this estimator. As Lee et al. (2020) point out this rule of thumb, while largely used, does not ensure that the standard t-statistics based inference is reliable. We think that the location of the structural parameters of the SVARs is the most important concern while naturally the related distributions are also important. Therefore, we utilize the more conventional procedures and base our inference on standard t-statistics as the efficient F-statistics is larger than 10. However, if we would use the Anderson and Rubin (1949) test instead, for example, the output elasticity of net revenue estimates we report are still significant at the 0.05 level.

We study the relevance assumption in more detail in the next section where we actually utilize our instrument in estimating SVAR models for both Canada and Finland. Here we show evidence related to the relevance assumption of our instrument that can be in a sense considered more general compared to the F-statistics from the first stage regressions. Table 2 (panel A) reports results from regressions of our proposed instrument on GDP forecast errors and for comparisons sake the Fernald (2014) TFP series for both the sample used in Caldara and Kamps (2017) and a more recent sample for USA. The forecast errors of output can be seen as proxying unexpected GDP movements. The results show that in all cases for Canada and Finland, our instrument is significantly related with unexpected GDP movements. This is also true for both TFP instruments and unexpected GDP movements in USA, which is the expected result, for example, based on the first stage statistics shown in Caldara and Kamps (2017).[15]

[14]Note that in the case where we have only one endogenous regressor and one instrument, this test reduces to the heteroskedasticity robust F-test.

[15]Additionally in appendix B in Table 6 (panel A) we also show results from a regressions were we regress

Exogeneity.— . The reasoning why our instrument should satisfy the exogeneity assumption is the following. Since the countries we consider are argued to be relatively small, unexpected shocks to their fiscal variables should not have a notable effect on their trading partners economic development contemporaneously. This arguably holds especially for a country with multiple trading partners. This is because, even if the unexpected fiscal shock is a direct government purchase (import) from a certain trading partner, the fact that the instrument is formed from unexpected changes in several trading partners, should dilute this relation.[16]

While the intuition behind the exogeneity assumption is credible given that we consider only relatively small economies, it would be convenient if this assumption could also be tested for statistically. Unfortunately, the exogeneity assumption cannot be straightforwardly tested. This is because we do not directly observe the exogenous shocks we are interested in. However, to test exogeneity, we gather one-period ahead forecast errors of government spending and investment from professional forecasts in the same manner as we do for the trading partner instrument. We then regress this proxy for the unexpected fiscal shock on the instrument with the VAR lags of endogenous variables as controls.[17]

In Panel B of Table 2 we show results from these regressions with different versions of our external instrument regressed on OECD forecast errors of government spending growth for Finland and Canada (columns 1-5). In the same table we also show results for the utilization-adjusted TFP series of Fernald (2014) in the case of the USA. We show the results using the series which is collected from Caldara and Kamps (2017) (column 6) and a more recent 2019 revision of the same utilization-adjusted TFP series (column 7) regressed against SPF forecast errors of government spending growth for the USA.[18] As these regressions show, the coefficients for Finland and Canada are all insignificant. This, we argue, supports our claim that the instrument we propose is indeed exogenous.

For the USA there is a statistically significant relation between Fernald's (2014) TFP series and the errors but not between the errors and the instrument Caldara and Kamps

the output residuals from our VAR against the different instruments. Accordingly, there exists a significant relation in each case.

[16]The exogeneity assumption also requires that an unexpected change in the trading partners economy does not result in an unexpected shock in government spending or revenue. Since the government does not see in real time whether a trading partner is doing better than expected it is unlikely that it would make such unexpected spending or revenue decision, which would be systematically related to the unexpected changes in trading partners economies.

[17]For example, Auerbach and Gorodnichenko (2012) use the OECD forecast errors of government spending and investment as proxies for fiscal shocks.

[18]Utilization-adjusted TFP growth of Fernald (2014) are from the November 14th 2019 vintage and downloaded from `https://sites.google.com/site/fernaldjg/TFP`. All other vintages used in this paper are from this source.

(2017) utilize. The utilization adjusted TFP series has seen many revisions over the years and this might according to Kurmann and Sims (forthcoming) have a considerable effect on estimation results. To ensure that the result in Table 2 Panel B column 7 considering Fernald's (2014) series is not just a one time wonder we show in appendix C in Figure 10 coefficient estimates and confidence intervals from the same regression using different vintages of Fernald's (2014) utilization adjusted TFP series. We find that overall the coefficient estimates are significant and positive across vintages. In appendix C we further study the relevancy and exogeneity of the Fernald's (2014) TFP series based instrument.

Another way to study the exogeneity assumptions is the following. We make the same assumption as in Blanchard and Perotti (2002) that government spending does not react within quarter to unexpected changes in output. In this paper we call this identification as BP_g. We can then test if the residual of government spending from the VAR has any predictive power on the instrument. These results are reported in appendix B in table 6 (panel B). Again we find evidence that our instrument is exogenous which is true also for the instrument Caldara and Kamps (2017) use. The TFP instrument based on Fernald (2014) TFP series is again maybe endogenous.[19]

[Table 2 here.]

5 Fiscal SVAR-IV

In this section we apply the instrument to estimating a simple fiscal SVAR model for both Canada and Finland. The analysis showcases the relevance of our proposed instrument. We also find that the choice between Blanchard and Perotti (2002) or Caldara and Kamps (2017) style identification has a marked effect on the estimated impulse responses. As we discuss later in more detail this difference is due to the additional zero restriction in Blanchard and Perotti (2002) identification rather than due to our instrument. The proposed instrument produces similar sized estimates of the output elasticity of net revenues in both identification schemes.

Our main specification of the reduced form VAR includes 5 lags and the endogenous variables are detrended before estimation. In tables 3 and 4 we report the estimated structural parameters related with this baseline reduced form specification.[20] The estimated size of the output elasticity of net revenue for Canada is 4.2-4.37 and for Finland 1.30-1.36

[19] One more detail considering Finland is that the relation between the instrument which bases on GDP forecast errors and unexpected government spending is significant at the 0.1 level. This suggests that maybe the imports based instrument is more likely exogenous than the GDP based one.

[20] Appendix E covers these tables more holistically.

depending on the identification scheme of the structural form. For Canada this estimate is notably larger than those shown in Perotti (2005) which are derived as in Blanchard and Perotti (2002).[21] For Finland the estimates are similar in size compared to earlier Blanchard and Perotti (2002) style estimates, see Lehmus (2014) and Virkola (2014).

[Table 3 here.]

[Table 4 here.]

5.1 Robustness of IV identification

How robust are the estimates for different specifications of the instrument and the SVAR? In Figures 2 and 3 we show the coefficient estimates and the F-statistics from the first stage regressions from several subsamples of the data for both Canada and Finland. The three columns each represent a different identification scheme. The first column represents Blanchard and Perotti (2002) style identification with the contemporaneous effect of structural revenue shocks on spending restricted to zero (BP_g) while the third column represents the corresponding Caldara and Kamps (2017) style identification (CK_g) where we estimate the contemporaneous elasticity of g to x instead of assuming it to be zero. The middle column represents both the Blanchard and Perotti (2002) style identification and the Caldara and Kamps (2017) style identification with the contemporaneous effect of structural government spending shocks on revenue restricted to zero (BP_r and CK_r), which in the case of output elasticity of net revenues, yield the same estimates.

In the upper panel of Figures 2 and 3 we show the coefficient estimates for output elasticity of net revenues. Each dot represents an estimate from a different subsample of the overall data. For example, in the case of Canada in Figure 2, the first dot represents the estimate from the subsample 1969Q1-2018Q4 and the next dot from 1970Q1-2018Q4 and so on with one year intervals. For Finland the first dot represents the period 1975Q1-2018Q4 the second 1976Q1-2018Q4 and so on. The shaded area represents the 0.95 robust confidence intervals for the estimates depicted with the black dots, which we consider as the baseline instrument in both Figures 2 and 3. In the lower panel of these figures we report the

[21] Perotti's (2005) estimate is 1.86. If we alternatively use the imports based instrument for Canada we get quite close to this. We anyway consider the GDP based instrument to be our baseline instrument. As we show later in this section also a specification with fiscal foresight variables produces an estimate, which is more close to the value Perotti (2005) reports. Furthermore, according to the results in Caldara and Camps (2017) SVAR-IV seems to produces notably larger values for this elasticity for USA, namely 3.6 (Mertens and Ravn (2011) unanticipated tax shocks instrument) and 2.4 (TFP instrument). Also Angelini, Caggiano, Castelnuovo and Fanelli (2020) find similar sized elasticities. In Perotti (2005) the value for USA is rather similar to that of Canada's.

first stage robust F-statistics for all of the corresponding coefficient estimates. A horizontal dotted line is drawn to indicate the conventional rule of thumb cutoff point of the robust F-statistic of 10.

In the figures we also consider different versions of the instrument. For Canada the black and orange lines represent instruments that are both based on GDP forecast errors with the distinction being that the black line represents the baseline case with the instrument constructed from SPF forecast errors of US real GDP while the orange line represents an instrument constructed from weighted OECD forecast errors of real GDP for all available trading partner countries. The blue line represents an instrument constructed from weighted OECD forecast errors of imports in the same manner as the OECD GDP instrument. For Finland we show results using four different specifications of the instrument, two based on weighted OECD GDP forecast errors and two based on weighted OECD import forecast errors. In the case of Finland, we show versions, for both the GDP and imports based instruments, where only G7 countries are used and a version where all available countries are used to form the instrument. For the G7 countries we have complete data for all years in the sample whereas for the other instrument additional countries are added whenever data are available.

Figure 2 shows results for Canada. While there is variation in the different estimates overall they are quite similarly sized and almost in all cases statistically significantly different from zero. The first stage F-statistic also suggests that none of the different versions of the instrument are weak and that the imports based instrument has a somewhat higher relevance than either of the GDP based instruments. Both of the GDP based instruments are near the threshold value of 10 in subsamples that start around the year 1990. Part of this could be explained by the sample size getting smaller as one moves leftwards in the figure. The import based instrument, however, seems to have higher relevance also with these subsamples while also after 1990 there is a clear decrease in the corresponding F-statistics. Interestingly, the imports based instrument also produces clearly a smaller elasticity estimate than either of the GDP based instruments.

Figure 3 shows corresponding results for Finland. Compared with the estimates for Canada the estimates for Finland are more stable over different subsamples. The estimates from samples starting from 1990 seem to be somewhat smaller than estimates from longer samples. Also, as in the case of Canada, the imports based instruments produce slightly smaller estimates than GDP based instruments in most subsamples. The robust F-statistic for the GDP instrument is above 10 in all samples which start before 1990. In samples starting from on 1990 the robust F-statistic is under 10 but still quite close to this reference value. The imports based instrument's F-statistics vary notably more than the GDP based instrument's F-statistics.

There are a couple of underlying issues to consider with regards to the relevance of our instrument. One issue potentially affecting the relevance of the instrument over time is that the share of exports to GDP has roughly doubled over our sample period in both Finland and Canada. In other words, both economies have become more open to trade over time. Also, a major part of Canada's exports go to USA while for Finland even the combination of G7 countries does not capture a similarly large share of exports. It might be that for Finland the imports based instrument starts gaining more relevance in the subsamples starting from the 1980s as additional countries are added to the OECD dataset and thus a larger share of trading partner countries are covered. The variation in the share of exports covered by the trading partners of Finland can be seen from Figure 17 in Appendix H. In Figure 3, the deviation between the two imports based instruments starting from somewhere around 1980 probably springs from this; more countries are available datawise and this makes the instrument more relevant in the later years of the overall sample. While this deviation is not as clear among the GDP based instruments it seems that the addition of extra countries slightly improves the instruments relevancy. Also it seems that after 1980 the F-statistics of the GDP based instruments and the imports based instruments are more close to each other. This might also reflect the increased weight of exports overall.

We have also looked at the robustness of the estimated elasticities to changes in the reduced form VAR. These results are discussed in more detail in Appendix F. Overall, neither lag length selection or the choice of using detrended or log-level data in the VAR affect our estimates in any significant way. Moreover, the instrument is relevant across different specifications.

[Figure 2 here.]

[Figure 3 here.]

5.2 Impulse responses

In Figures 4 and 5 we depict the impulse response for Canada and for Finland using different identifications of the structural form. For both countries all the impulse responses that relate to the set $\{r, x\}$ are very similar across identifications. In contrast, impulse responses that are related to g vary to different degrees between the identifications. Crucially, when considering fiscal multipliers the response from government spending to output is the primary interest.

These differences between the identifications seem to mostly stem from the additional zero restriction inherent in BP identification. This finding causes a predicament. The BP style zero restriction of the government spending response to output within a quarter is well established in the literature. At the same time, the CK style identification does not produce statistically significant estimates of this coefficient (expect for Finland with CK_r identification at the 0.1 significance level).[22] It seems reasonable that one could just set this parameter zero. However, whether this parameter is zero or not has a crucial impact on the impulse responses of main interest. Therefore, while the parameter might be statistically insignificant, the effect on the impulse responses is potentially a drastic one.[23]

[Figure 4 here.]

[Figure 5 here.]

Figure 6 showcases the importance of the output elasticity estimates and the effect of the BP zero restriction on the contemporaneous impact multipliers of fiscal variables. In this figure we depict impact multipliers (y-axis) as a function of the output elasticities of government spending or net revenues (x-axis) as well as the estimates from different identifications as colored dots on the resulting curves.[24] Output elasticity is a key parameter as it effectively determines how much of the contemporaneous variation in the fiscal variable can be explained by changes in output. This in turn then largely determines how much of the variation in output can be explained by the fiscal variable in a SVAR. Given that the underlying reduced form VAR is fixed, differences in the contemporaneous responses stemming from the specification of the structural form then explain much of the differences seen in impulse responses. A more detailed discussion on this topic can be found in Caldara and Kamps (2017) where a similar exercise is carried out.

Panels a and c in Figure 6 illustrate how sensitive the government spending impact multiplier is to changes in the output elasticity of g. For Canada the estimated output elasticity (red dotted line) is negative and this results in a higher impact multiplier than the

[22] For Finland with CK_r identification this parameter is not significant at the 0.1 significance level if Anderson and Rubin (1949) test is used instead of standard t-statistic based inference.

[23] It could be that while this estimate does not seem to be statistically significant in a finite sample, it still might not be zero. Furthermore, if this parameter is not set to zero, it also has an effect on the size of the estimate of the contemporaneous response of output to government spending (row g column x in Tables 3 and 4). For Finland this estimate becomes more insignificant when changing from BP to CK whereas for Canada this estimate becomes significant. Furthermore, for Finland the sign of this parameter is negative with CK identification which is somewhat unintuitive as one would expect that contemporaneously if anything output reacts positively on government spending. For Canada the sign is positive and the estimate is larger with CK than with BP identification.

[24] We set the parameter value on the x-axis outside the model and estimate all the other parameters in the usual manner using CK style identification. Note that across this paper BP identification can be seen as just a special case of the CK identification where the output elasticity of government spending and investment is set to zero and thus we find BP identification naturally as a point on the curve.

BP zero restriction yields. For CK_g the impact multiplier is roughly unity while for BP_r it is close to zero. However, in all cases the impact multiplier is positive. In contrast, for Finland the estimated output elasticity of g is positive and this results in a negative impact multiplier while the BP zero restriction yields a positive one. As stated before, we cannot reject the BP zero restriction at the 0.05 level as for both countries the 0.95 confidence interval (shaded region) contains zero.

In line with our preceding analysis, we see from the colored dots in panels b and d of Figure 6 that on the net revenue side there are no large differences in the output elasticity estimates between different identifications. It is also notable that even though the elasticity estimate for Canada is much larger than it is for Finland, the impact multipliers are not that different. Note also that for Finland the difference in impact multipliers associated with both ends of the confidence interval for the elasticity estimate is actually larger even though the confidence interval is much narrower than it is for Canada.

[Figure 6 here.]

Evidently also the reduced form VAR specification has an effect on the impulse responses and thereby estimates of the fiscal multiplier. In Figures 7 and 8 we plot the impulse responses of the SVAR model identified with the four different schemes using different lag lengths in the reduced form VAR. Looking at the figures, it seems that while the lag order has a clear effect on some of the impulse responses, different lag length specifications still produce rather similar impulse responses overall. Impulse responses using different lag lengths mostly fall on the 0.68 confidence interval of the baseline specification. Differences in impulses responses seem to be driven more by the choice of identification scheme than by changes to the reduced form VAR specification.

[Figure 7 here.]

[Figure 8 here.]

5.3 Fiscal foresight

Arguably a 3-variable fiscal VAR is not rich enough in information to uncover the true structural shocks in a credible way whichever identification strategy the researcher uses. All VARs assume that all endogenous variation in model variables is explained by p lags of endogenous variables. In the identification of fiscal shocks, fiscal foresight plays a potentially major role and a number of recent papers have shown how fiscal SVARs that do not account for fiscal foresight might be misspecified. Ramey (2011) shows how past forecasts and announcements predict VAR shocks and how taking this foresight into account affects estimates of the government spending multiplier. Leeper, Walker and Yang (2013) and Forni and Gambetti (2014) discuss how the econometric analysis is generally complicated by the presence of foresight and the conditions of informational sufficiency under which the structural shocks can be recovered respectively.

To study how fiscal foresight affects our SVAR-IV identification and the resulting impulse responses, we augment our three variable model by extending the vector of endogenous variables to include variables that account for foresight in the spirit of Forni and Gambetti (2016). Specifically, we collect real-time forecast revisions of government expenditures and GDP from the Bank of Canada staff economic projections for 1987Q1-2013Q4 (Champagne et al., 2018).[25] Unfortunately, quarterly forecasts for Finland are not available for this long a period and thus our focus here is on Canada.

From forecast revisions we construct a variable that aims to capture news shocks in the VAR system. Following Forni and Gambetti (2016), we build this variable as a cumulative sum of the forecast revisions. Let n_t^j be a variable that is the cumulative sum of forecast revisions of j up to time t. We augment the original 3-variable VAR by including this variable as an endogenous variable.

As a new variable is included in the reduced form VAR, we need to reconsider also the identification of the structural form. Since we have an instrument for output, it makes sense to estimate the elasticity of contemporaneous n_t^j to x_t via 2SLS as is done with the other endogenous variables. With respect to g and r we assume that the news shock (the structural shock to n^j) affects these variables only through its effect on x. In effect we thus augment the identification by running an additional 2SLS regresssion after e^g and e^r have been identified in which we find coefficients for the aforementioned structural shocks as well as the elasticity of contemporaneous n_t^j to x_t. The identification strategy of e^g and e^r thus remains unchanged in all our identification schemes but of course the reduced form residuals are changed under the presence of an additional variable in the system.

We are interested in how fiscal foresight might affect our instrumental identification strategy. To this effect we augment our VAR with foresight variables for both government expenditure and output, separately and in tandem. To see how the foresight variables affects our estimated coefficients we provide some additional results that can be compared to our previous analysis. In Table 5 We call these specifications Foresight g, Foresight x and Foresight g and x, while at the same time the 3-variable VAR is called the Baseline specification.

The 2SLS regressions where we explain contemporanous net revenues are presented in

[25]These data are available at `https://www.bankofcanada.ca/rates/staff-economic-projections/`.

Table 5. Note that to add comparability, the sample period is the same (1987Q1-2013Q4) in all specifications and also all regressions have 5 lags of endogenous variables as controls. It can be seen that without the foresight variables, especially the foresight variable for output, the strength of the instrument is weaker and at the same time the estimated elasticity of net revenues to output is larger across all identification schemes. When both foresight variables are included, the estimated elasticity of net revenues to output is much closer to traditional elasticities that have been reported earlier in the literature (see Blanchard and Perotti (2002) and Perotti (2005)).

Figure 9 shows how the adding of foresight variables to the VAR model effects on the estimated impulse responses. The first column consists of the impulse responses from a unit government spending shock to output while the second column plots the impulse responses from unit net revenue shock to output. Each row represents a different reduced form VAR and different identification schemes are marked by different colored lines. It is evident that different identification schemes produce different impulse responses from g to x even in the presence of foresight variables. At the same time impulse responses from r to x do not vary much across specifications. Between the different foresight models it seems that adding the foresight variable for x does not have a large effect on the impulse responses when compared to the baseline specification while adding the foresight variable for g has a small downward effect on the estimated impulse responses from g to x.

[Table 5 here.]

[Figure 9 here.]

5.4 Fiscal multipliers

We calculate cumulative fiscal multipliers using each of the four identification schemes for both countries. Here we briefly review our main findings while a more detailed discussion is relegated to Appendix G. We want to emphasise that one should take these estimates with a grain of salt. To achieve a reliable and robust fiscal multiplier estimate, more care should arguably be taken in building a more credible model.[26]

For Canada we find cumulative government spending multipliers ranging from -0.83 to 0.90 over one year horizon and from -0.94 to 0.90 over two year horizon. The estimates are mostly insignificantly different from zero at the 0.68 level. A near zero multiplier would

[26]The 3-variate SVAR model is probably not rich enough in information to produce such an estimate. Moreover, the exchange rate regime shift in the case of Finland seems to play an important role. In addition, rare events such as the deep depression in the beginning of the 1990s in Finland perhaps have an outsized effect on the estimates and thus warrant special consideration.

be consistent with the prediction of the canonical Mundell-Fleming model for a small open economy with a floating exchange rate.

In the case of Finland we find cumulative government spending multipliers ranging from -0.55 to 2.03 over one year horizon and from -0.59 to 4.96 over two year horizon. The multipliers using CK identification are systematically smaller than multipliers found using BP identification. When estimating the multiplier using a sample starting from 1995Q2, which mostly covers the EMU period, we find notably larger multipliers than using the whole sample. This could be since for a small country like Finland in a large monetary union, monetary policy is practically exogenous and the exchange rate effectively fixed.

6 Conclusion

In this paper we have proposed a novel instrument for aggregate output that is based on professional forecast errors in trading partner economies. The instrument we propose is suitable for a small open economy setting where the domestic economy has a negligible effect on the economies of its trading partners while at the same time the developments in these trading partners are highly relevant for the domestic economy.

This instrument, we argue, has a number of desirable properties when compared to the prevailing instrument for output used in the related literature which is the utilization-adjusted TFP series of Fernald (2014). Our instrument is quite simple to construct from observable data and there is arguably a much smaller role for the researcher in making choices that affect the resulting series. We provide evidence for the validity of our proposed instrument. Test results show the instrument to be relevant and not weak. More so, we also find suggestive evidence that the instrument seems to fulfill the necessary exogeneity assumption. In comparison we raise some questions about the exogeneity of the utilization-adjusted TFP series with respect to government spending shocks. Furthermore, we show that revisions made to the TFP series have had a potentially large effect on its performance as an instrument. A disadvantage of our proposed instrument is that it is valid only for small open economies.

We apply the proposed instrument in estimating fiscal SVAR models using the SVAR-IV methodology for Canada and Finland. As a result of the fiscal SVAR analysis we get estimates for impulse responses and corresponding fiscal multipliers. We find that some of the impulse responses as well as government spending multipliers are sensitive to differences in SVAR identification schemes we consider (Blanchard and Perotti (2002) and Caldara and Kamps (2017)) However, broadly taken, the fiscal multipliers we calculate are in line with what one would expect in a small open economy context.

Our main emphasis in this paper is on validating our instrument. We stress that more care should be taken in specifying the underlying model to produce more credible fiscal multiplier estimates. For example, as acknowledged in the literature, it is questionable whether a 3-variable VAR is rich enough in information. Therefore, foresight variables that contain relevant information in predicting changes in fiscal policy can be crucial for the model to correctly identify actual policy shocks. However, also a number of other variables might be relevant and potentially added to the model. We want to point out that there is no reason why our instrument cannot be used together with other instruments to identify a more complex SVAR model. We also think that methods where the instrument enters the SVAR model internally like in Noh (2018) or in Angelini et al. (2020) might be preferable to estimating the structural parameters with external instruments as in our SVAR-IV application in the manner of Stock and Watson (2018).

Finally, we want to emphasize that the instrument we propose can potentially be used also in other contexts. There is no reason why one would be limited to utilize our instrument in estimating fiscal multipliers in a SVAR model. As an example, our instrument and the structural identifications we utilize can also be applied in local projections, see Plagborg-Møller and Wolf (forthcoming). Presumably there are many other potential applications where this instrument might be useful.

References

Anderson, T. W., Rubin, H. et al.: 1949, Estimation of the parameters of a single equation in a complete system of stochastic equations, *The Annals of Mathematical Statistics* **20**(1), 46–63.

Andrews, I., Stock, J. H. and Sun, L.: 2019, Weak instruments in instrumental variables regression: Theory and practice, *Annual Review of Economics* **11**, 727–753.

Angelini, G., Caggiano, G., Castelnuovo, E. and Fanelli, L.: 2020, Are fiscal multipliers estimated with proxy-SVARs robust?, *Quaderni-Working Paper DSE* .

Auerbach, A. J. and Gorodnichenko, Y.: 2012, Fiscal multipliers in recession and expansion, *Fiscal policy after the financial crisis*, University of Chicago Press, pp. 63–98.

Awokuse, T. O.: 2003, Is the export-led growth hypothesis valid for Canada?, *Canadian Journal of Economics* **36**(1), 126–136.

Basu, S., Fernald, J. G. and Kimball, M. S.: 2006, Are technology improvements contractionary?, *American Economic Review* **96**(5), 1418–1448.

Blanchard, O.: 2019, Public debt and low interest rates, *American Economic Review* **109**(4), 1197–1229.

Blanchard, O. J. and Summers, L. H.: 2017, Rethinking stabilization policy: Evolution or revolution?, *National Bureau of Economic Research Working Paper* .

Blanchard, O. and Perotti, R.: 2002, An empirical characterization of the dynamic effects of changes in government spending and taxes on output, *Quarterly Journal of Economics* **117**(4), 1329–1368.

Brüggemann, R., Jentsch, C. and Trenkler, C.: 2016, Inference in VARs with conditional heteroskedasticity of unknown form, *Journal of Econometrics* **191**(1), 69–85.

Caldara, D. and Kamps, C.: 2017, The analytics of SVARs: a unified framework to measure fiscal multipliers, *The Review of Economic Studies* **84**(3), 1015–1040.

Champagne, J., Bellisle, G.-P. and Sekkel, R.: 2018, Evaluating the Bank of Canada staff economic projections using a new database of real-time data and forecasts.

Comin, D. A., Gonzalez, J. Q., Schmitz, T. G. and Trigari, A.: 2020, Measuring TFP: The Role of Profits, Adjustment Costs, and Capacity Utilization, *Working Paper 28008*, National Bureau of Economic Research.

Dorich, J., Johnston, M. K., Mendes, R. R., Murchison, S. and Zhang, Y.: 2013, ToTEM II: An updated version of the Bank of Canada's quarterly projection model.

Efron, B.: 1979, Bootstrap methods: Another look at the jackknife, *Annals of Statistics* (7), 1–26.

Efron, B.: 1982, *The jackknife, the bootstrap, and other resampling plans*, Vol. 38, SIAM.

Fernald, J. G.: 2014, A quarterly, utilization-adjusted series on total factor productivity, *Federal Reserve Bank of San Francisco Working paper Series* **2012-19**.

Forni, M. and Gambetti, L.: 2014, Sufficient information in structural VARs, *Journal of Monetary Economics* **66**, 124–136.

Forni, M. and Gambetti, L.: 2016, Government spending shocks in open economy VARs, *Journal of International Economics* **99**, 68–84.

Gorodnichenko, Y., Mendoza, E. G. and Tesar, L. L.: 2012, The Finnish Great Depression: From Russia with love, *American Economic Review* **102**(4), 1619–1644.

Gulan, A., Haavio, M. and Kilponen, J.: 2019, Can large trade shocks cause crises? The case of the Finnish-Soviet trade collapse, *Bank of Finland Research Discussion Paper* (9).

Hall, P.: 1992, *The bootstrap and Edgeworth expansion*, Springer-Verlag, New York.

Ilzetzki, E., Mendoza, E. G. and Végh, C. A.: 2013, How big (small?) are fiscal multipliers?, *Journal of monetary economics* **60**(2), 239–254.

Jentsch, C. and Lunsford, K. G.: 2019, The dynamic effects of personal and corporate income tax changes in the United States: Comment, *American Economic Review* **109**(7), 2655–78.

Kilian, L.: 1998, Small-sample confidence intervals for impulse response functions, *Review of economics and statistics* **80**(2), 218–230.

Kilian, L.: 1999, Finite-sample properties of percentile and percentile-t bootstrap confidence intervals for impulse responses, *Review of Economics and Statistics* **81**(4), 652–660.

Kilian, L. and Lütkepohl, H.: 2017, *Structural vector autoregressive analysis*, Cambridge University Press.

Kurmann, A. and Sims, E.: forthcoming, Revisions in utilization-adjusted TFP and robust identification of news shocks, *Review of Economics and Statistics* .

Lee, D. L., McCrary, J., Moreira, M. J. and Porter, J.: 2020, Valid t-ratio inference for IV.

Leeper, E. M., Walker, T. B. and Yang, S.-C. S.: 2013, Fiscal foresight and information flows, *Econometrica* **81**(3), 1115–1145.

Lehmus, M.: 2014, Finnish fiscal multipliers with a structural VAR model, *Labor Institute for Economic Research working paper* **293**.

Mertens, K. and Ravn, M. O.: 2011, Understanding the aggregate effects of anticipated and unanticipated tax policy shocks, *Review of Economic dynamics* **14**(1), 27–54.

Mertens, K. and Ravn, M. O.: 2013, The dynamic effects of personal and corporate income tax changes in the United States, *American Economic Review* **103**(4), 1212–47.

Mertens, K. and Ravn, M. O.: 2019, The dynamic effects of personal and corporate income tax changes in the United States: Reply, *American Economic Review* **109**(7), 2679–91.

Montiel Olea, J. L. and Pflueger, C.: 2013, A robust test for weak instruments, *Journal of Business & Economic Statistics* **31**(3), 358–369.

Mountford, A. and Uhlig, H.: 2009, What are the effects of fiscal policy shocks?, *Journal of applied econometrics* **24**(6), 960–992.

Nakamura, E. and Steinsson, J.: 2018, Identification in macroeconomics, *Journal of Economic Perspectives* **32**(3), 59–86.

Noh, E.: 2018, Impulse-response analysis with proxy variables, *Manuscript, University of California San Diego* .

Perotti, R.: 2005, Estimating the effects of fiscal policy in OECD countries, *IGIER working paper* **276**.

Plagborg-Møller, M. and Wolf, C. K.: forthcoming, Local projections and VARs estimate the same impulse responses, *Econometrica* .

Ramey, V. A.: 2011, Identifying government spending shocks: It's all in the timing, *The Quarterly Journal of Economics* **126**(1), 1–50.

Ramey, V. A.: 2019, Ten years after the financial crisis: What have we learned from the renaissance in fiscal research?, *Journal of Economic Perspectives* **33**(2), 89–114.

Ramey, V. A. and Zubairy, S.: 2018, Government spending multipliers in good times and in bad: evidence from us historical data, *Journal of Political Economy* **126**(2), 850–901.

Romer, C. D. and Romer, D. H.: 2010, The macroeconomic effects of tax changes: estimates based on a new measure of fiscal shocks, *American Economic Review* **100**(3), 763–801.

Solow, R. M.: 1957, Technical change and the aggregate production function, *The review of Economics and Statistics* pp. 312–320.

Stock, J. H. and Watson, M. W.: 2018, Identification and estimation of dynamic causal effects in macroeconomics using external instruments, *The Economic Journal* **128**(610), 917–948.

Stock, J. and Yogo, M.: 2005, *Testing for Weak Instruments in Linear IV Regression. In: Andrews DWK Identification and Inference for Econometric Models*, Cambridge University Press, New York.

Virkola, T.: 2014, Exchange rate regime, fiscal foresight and the effectiveness of fiscal policy in a small open economy, *ETLA Report 20* .

Tables

Identification	a_{gx}	b_{gr}	a_{rx}	b_{rg}	a_{xg}	a_{xr}
BP_g	0	0	$\checkmark(m_x)$	$\checkmark$	$\checkmark(\widehat{e_g})$	$\checkmark(\widehat{e_r})$
BP_r	0	$\checkmark$	$\checkmark(m_x)$	0	$\checkmark(\widehat{e_g})$	$\checkmark(\widehat{e_r})$
CK_g	$\checkmark(m_x)$	0	$\checkmark(m_x)$	$\checkmark$	$\checkmark(\widehat{e_g})$	$\checkmark(\widehat{e_r})$
CK_r	$\checkmark(m_x)$	$\checkmark$	$\checkmark(m_x)$	0	$\checkmark(\widehat{e_g})$	$\checkmark(\widehat{e_r})$

Table 1: SVAR identification schemes

Notes: This table illustrates the differences between the 4 SVAR identification schemes utilized across the paper. Each row represents one of these 4 schemes. Each column represents one of the structural parameters. Zero implies that the parameter is restricted to zero and a checkmark implies that the parameter is estimated. For those parameters which are estimated with 2SLS, the instrument that is used in estimation is given in brackets. In concrete terms, m_x stands for the proposed trading partner forecast error instrument which is used to instrument output x when estimating parameters a_{gx} and a_{rx}. On the other hand, $\widehat{\epsilon_g}$ and $\widehat{\epsilon_r}$ are the estimated exogenous series which are derived from prior steps before the estimation of parameters a_{xg} and a_{xr}. BP_g and BP_r refers to Blanchard and Perotti (2002) identification with government spending first (g) and with net revenue first (r) respectively. CK_g and CK_r refers to Caldara and Kamps (2017) style identification with government spending first (g) and with net revenue first (r) respectively.

Panel A: Relevance

	Dependent variable: Forecast error of output						
	CAN	CAN	CAN	FIN	FIN	USA (CK)	USA
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
GDP (SPF)	0.32^{***}						
	(0.08)						
GDP (OECD)		0.37^{***}		1.66^{***}			
		(0.09)		(0.56)			
Imports (OECD)			0.11^{***}		0.49^{***}		
			(0.02)		(0.17)		
Util. adj. TFP						0.43^{***}	0.28^{***}
						(0.10)	(0.07)
N	200	176	176	88	88	152	188
Adjusted R^2	0.23	0.26	0.28	0.44	0.40	0.15	0.11

Panel B: Exogeneity

	Dependent variable: Instrument						
	CAN			FIN		USA	
	GDP (SPF)	GDP (OECD)	Imports (OECD)	GDP (OECD)	Imports (OECD)	Util. adj. TFP (CK)	Util. adj. TFP
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
FE of g (OECD)	0.01	-0.03	-0.18	0.04	0.01		
	(0.07)	(0.06)	(0.20)	(0.03)	(0.11)		
FE of g (SPF)						-0.02	0.28^{***}
						(0.10)	(0.09)
N	132	132	132	88	88	101	137
Adjusted R^2	0.14	0.12	0.21	0.23	0.36	0.16	0.12

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table 2: Relevance and exogeneity of the instrument

Notes: In Panel A each column reports OLS estimation results from a regression of professional forecast errors of quarterly growth of GDP on different instruments. For Canada and Finland forecast errors of GDP are from OECD Economic Outlooks and for the US they are from the Survey of Professional forecasters (SPF). Instrument GDP (SPF) is constructed from US GDP forecast errors (SPF) and GDP (OECD) and Import (OECD) instruments from a export-share weighted mean of OECD trading partner forecast errors (OECD). Utilization-adjusted TFP series of Fernald (2014) is either from Caldara and Kamps or from the November 14th 2019 vintage and downloaded from `https://sites.google.com/site/fernaldjg/TFP`*. In Panel B, each column reports OLS estimation results from a regression of different instrument (now the dependent variable) on the professional forecast errors in the growth of the sum of general government consumption and investment. For Canada and Finland forecast errors of government spending are from the OECD Economic Outlooks and for the US they are from the Survey of Professional forecasters (SPF). In both panels, each model includes lags of VAR endogenous variables as controls; 5 lags for both Canada and Finland and 4 lags for the USA. Robust Eicker–Huber–White standard errors are used throughout.*

	BP_g		BP_r			CK_g			CK_r		
	r	x	r	g	x	g	r	x	r	g	x
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
x	4.20***		4.37***			−0.27	4.37***		4.37***	−0.27	
	(1.19)		(1.23)			(0.27)	(1.20)		(1.23)	(0.26)	
$\widehat{e^g}$	−0.65*						−1.09**				
	(0.39)						(0.43)				
$\widehat{e^r}$				−0.04***						−0.06*	
				(0.02)						(0.03)	
g		0.07			0.07			0.26***			0.26***
		(0.09)			(0.09)			(0.08)			(0.08)
r		−0.19***			−0.20***			−0.19***			−0.19***
		(0.04)			(0.05)			(0.05)			(0.05)
N	200	227	200	227	227	200	200	227	200	200	227
1st stage F-stat. (x)	20.22		19.22			19.22	20.54		19.22	34.67	
1st stage F-stat. (r)		46			41			45			45
1st stage F-stat. (g)		>999			>999			>999			>999
Wu-Hausman p-val.	0.00	0.00	0.00		0.00	0.11	0.00	0.00	0.00	0.23	0.00
Adjusted R^2	0.85	0.96	0.85	0.99	0.96	0.98	0.86	0.96	0.85	0.98	0.96

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 3: Structural parameter estimates, Canada

Notes: Each column represents estimation results for SVAR structural parameters across the different identification schemes. BP_g (columns 1-2) and BP_r (columns 3-5) refers to Blanchard and Perotti (2002) identification with government spending first (g) and with net revenue first (r) respectively. CK_g (columns 6-8) and CK_r (columns 9-11) refers to Caldara and Kamps (2017) style identification with government spending first (g) and with net revenue first (r) respectively. All of the above models contain these 5 lags of (detrended) endogenous variables as controls as well as the variables of interest shown in the first column. All columns represent a 2SLS-model except column (4) where the model is a standard OLS since there are no endogenous regressors. Output x is instrumented in each case with US GDP forecast errors and g and r with the estimated structural shocks $\widehat{e^g}$ and $\widehat{e^r}$ as identified within the identification scheme. For each identification scheme the order of columns represents the sequential order of estimation steps. Robust Eicker–Huber–White standard errors are used throughout.

	BP_g		BP_r			CK_g			CK_r		
	r	x	r	g	x	g	r	x	r	g	x
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
x	1.30**		1.36**			0.25	1.36***		1.36**	0.25*	
	(0.53)		(0.53)			(0.16)	(0.51)		(0.53)	(0.14)	
$\widehat{e^g}$	0.23*						0.33**				
	(0.13)						(0.14)				
$\widehat{e^r}$				0.06						0.08**	
				(0.03)						(0.04)	
g		0.13*			0.13*			−0.12			−0.12
		(0.07)			(0.07)			(0.09)			(0.09)
r		−0.12			−0.14			−0.12			−0.12
		(0.08)			(0.09)			(0.08)			(0.08)
N	171	171	171	171	171	171	171	171	171	171	171
1st stage F-stat. (x)	16.16		17.21			17.21	18.62		17.21	24.18	
1st stage F-stat. (r)		67			59			61			61
1st stage F-stat. (g)		>999			>999			>999			>999
Wu-Hausman p-val.	0.14	0.00	0.11		0.00	0.16	0.14	0.00	0.11	0.27	0.00
Adjusted R^2	0.97	0.96	0.97	0.94	0.96	0.94	0.97	0.96	0.97	0.94	0.96

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 4: Structural parameter estimates, Finland

Notes: Each column represents estimation results for SVAR structural parameters across the different identification schemes. BP_g (columns 1-2) and BP_r (columns 3-5) refers to Blanchard and Perotti (2002) identification with government spending first (g) and with net revenue first (r) respectively. CK_g (columns 6-8) and CK_r (columns 9-11) refers to Caldara and Kamps (2017) style identification with government spending first (g) and with net revenue first (r) respectively. All of the above models contain these 5 lags of (detrended) endogenous variables as controls as well as the variables of interest shown in the first column. All columns represent a 2SLS-model except column (4) where the model is a standard OLS since there are no endogenous regressors. Output x is instrumented in each case with weighted trading partner GDP forecast errors and g and r with the estimated structural shocks $\widehat{e^g}$ and $\widehat{e^r}$ as identified within the identification scheme. For each identification scheme the order of columns represents the sequential order of estimation steps. Robust Eicker–Huber–White standard errors are used throughout.

	Dependent variable: r											
	Baseline			Foresight g			Foresight x			Foresight g and x		
	BP_g	CK_g	BP_r/CK_r	BP_g	CK_g	BP_r/CK_r	BP_g	CK_g	BP_r/CK_r	BP_g	CK_g	BP_r/CK_r
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
x	3.74	3.32	3.32	2.91	2.54	2.54	2.76	2.33	2.33	2.14	1.80	1.80
	(2.52)	(2.27)	(2.14)	(2.27)	(2.05)	(2.00)	(1.70)	(1.57)	(1.54)	(1.63)	(1.53)	(1.51)
$\widehat{e}_g$	-1.53^{*}	-1.02		-1.06	-0.52		-1.43^{**}	-1.15^{**}		-1.14^{*}	-0.93	
	(0.87)	(0.68)		(0.86)	(0.66)		(0.61)	(0.54)		(0.68)	(0.62)	
N	103	103	103	103	103	103	103	103	103	103	103	103
1st stage F-stat. (x)	4.73	5.46	5.73	5.68	6.91	6.72	12.31	13.35	14.02	12.39	13.53	13.80
Wu-Hausman p-val.	0.08	0.08	0.08	0.21	0.21	0.23	0.21	0.21	0.18	0.46	0.46	0.45
Adjusted R^2	0.85	0.85	0.85	0.87	0.87	0.87	0.88	0.88	0.87	0.88	0.88	0.88

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table 5: Output elasticity estimates for Canada w/ foresight, sample 1987-2013

Notes: This table reports the coefficient estimates of two structural parameters (standard errors are given brackets) concerning the VAR specifications which include the foresight variables. The first parameter is the output elasticity of net revenue (x) and the second one is the relation between the government spending shock and the net revenue ($\widehat{e}_g$). Columns 1-3 report results from the 4 SVAR-IV identification schemes for the baseline specification where no foresight variables are included. Columns 4-6 contain results from a VAR specification where a government spending foresight variable is included and columns 7-9 contain results from a VAR specification where a output foresight variable included. Columns 10-12 contain results from a specification with both government spending and output foresight variables included. The reported first stage F-statistic is efficient. All specifications are with detrended data and include 5 lags. The estimation sample covers the years 1987-2013 for all specifications. BP_g and BP_r refers to Blanchard and Perotti (2002) identification with government spending first (g) and with net revenue first (r) respectively. CK_g and CK_r refers to Caldara and Kamps (2017) style identification with government spending first (g) and with net revenue first (r) respectively.

Figures

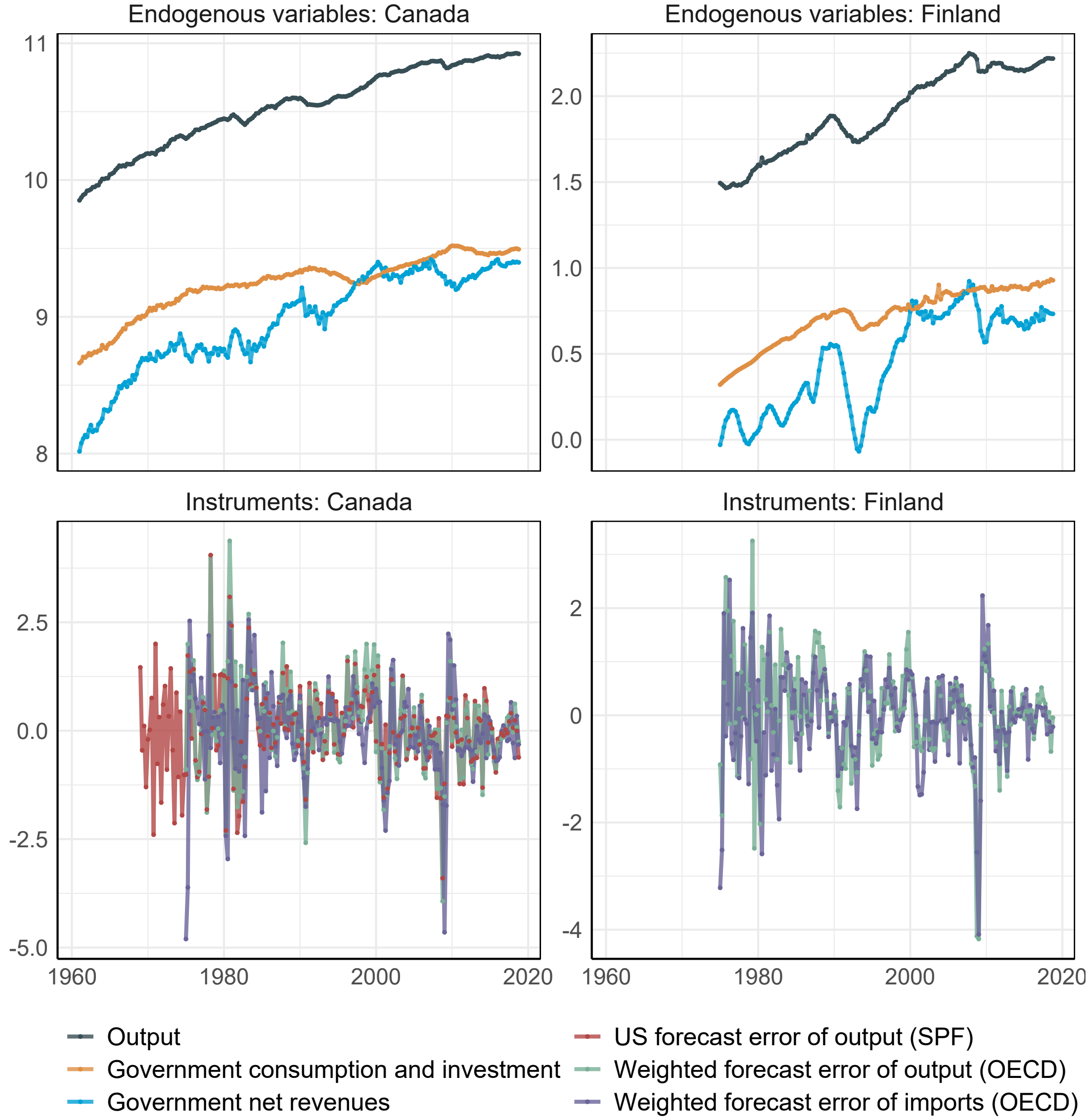

Figure 1: Data for the endogenous variables and the instruments.

Notes: Output (GDP), government consumption and investment as well as government net revenues in the upper panels are all in real per capita terms and in natural logarithms. The source for these data are the official statistical agencies of Canada and Finland. Forecast errors are calculated as the difference between the one period ahead log-difference in the level forecast and the corresponding realized log-difference in ex post data. US forecast errors are from the Survey of Professional forecasters while the weighted forecast errors for both output and imports are from past vintages of the OECD Economic Outlook. In the latter series, forecast errors for all available countries are weighted by their relative share in the exports of the country of interest. For illustrative purposes, all of the instruments presented in the lower panels have been scaled by their respective standard deviations.

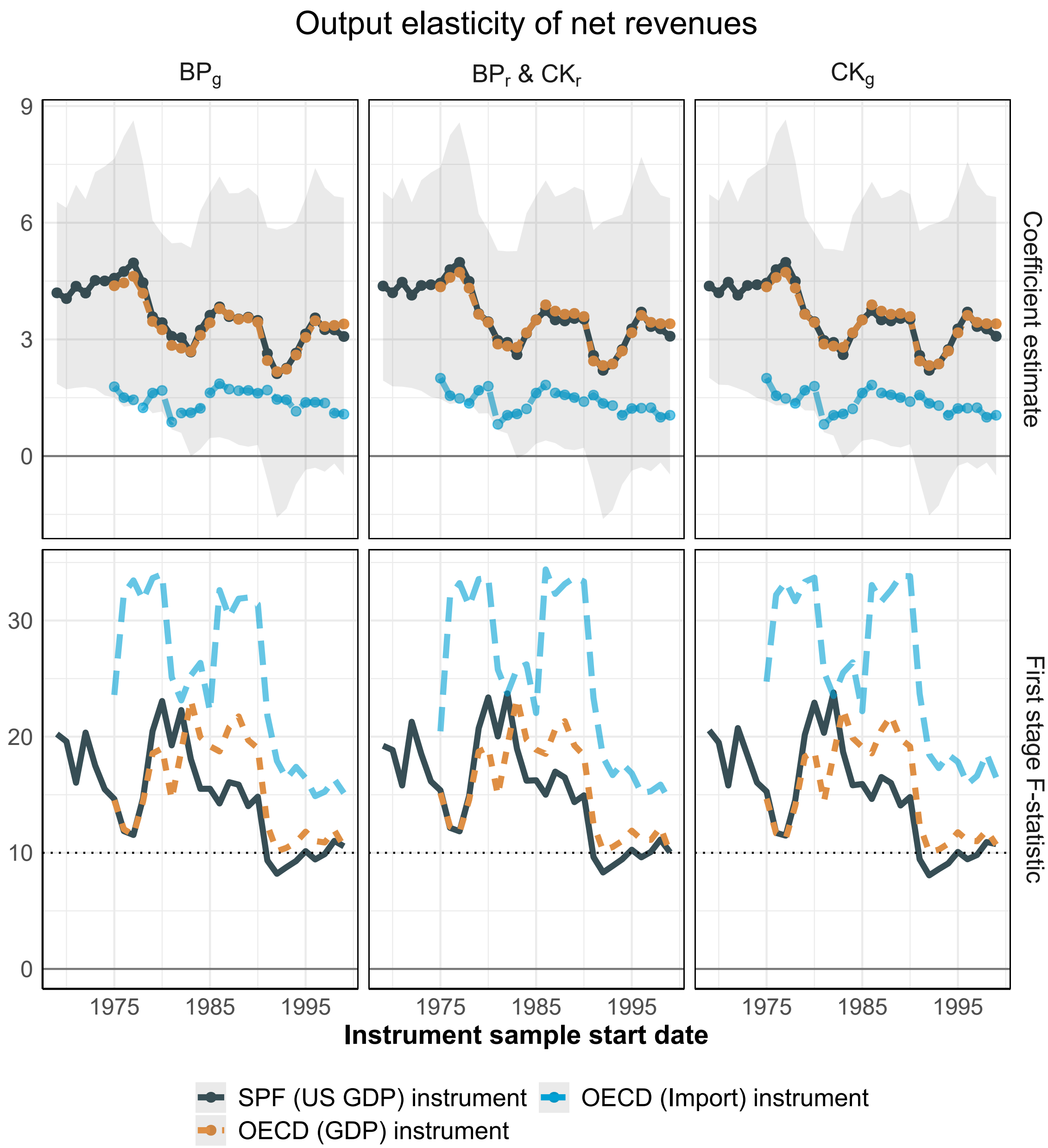

Figure 2: 2SLS estimates of the output elasticity of net revenues and robust 1st stage statistics over different sample periods and using different instruments, Canada

Notes: Each line represents the estimated coefficient and the corresponding first stage F-statistic from the estimation of output elasticity of net revenues. The x-axis represents the instrument sample start date starting from 1969Q1 that is used in each estimation (one year intervals). The left-most estimate is for the full-sample (1969Q1-2018Q4) while the rightmost estimate uses data only for 1999Q1-2018Q4. The shaded area represents the 0.95 robust confidence intervals for the estimate that uses the instrument that is constructed from forecast errors of all sample countries (black line). VAR specification has 5 lags and endogenous variables have been detrended prior to estimation. BP_g *and* BP_r *refers to Blanchard and Perotti (2002) identification with government spending first (g) and with net revenue first (r) respectively.* CK_g *and* CK_r *refers to Caldara and Kamps (2017) style identification with government spending first (g) and with net revenue first (r) respectively.*

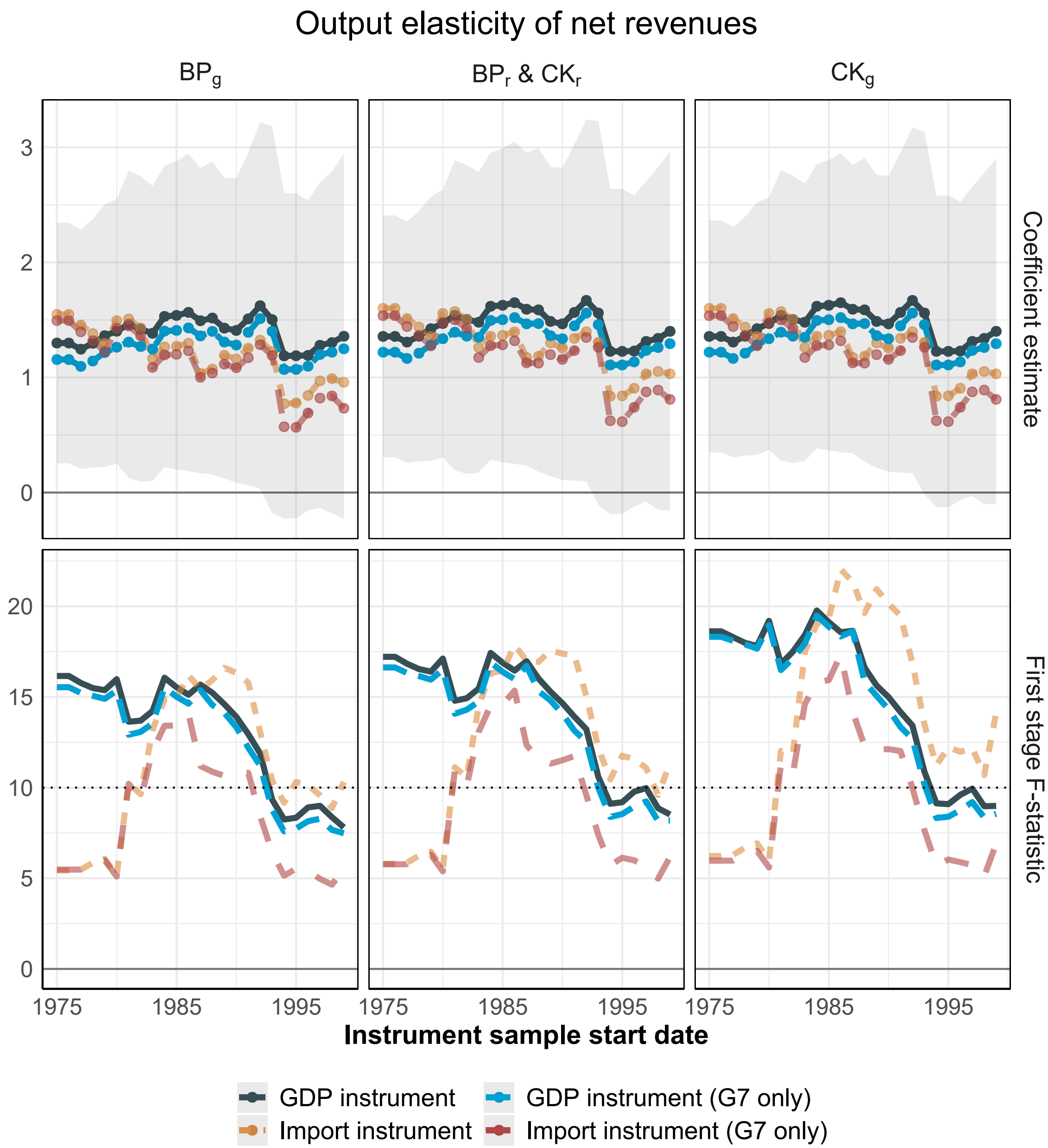

Figure 3: 2SLS estimates of the output elasticity of net revenues and robust 1st stage statistics over different sample periods and using different instruments, Finland

Notes: Each line represents the estimated coefficient and the corresponding first stage F-statistic from the estimation of output elasticity of net revenues. The x-axis represents the instrument sample start date starting from 1975Q1 that is used in each estimation (one year intervals). The left-most estimate is for the full-sample (1975Q1-2018Q4) while the rightmost estimate uses data only for the time that Finland has been part of the European monetary union (1999Q1-2018Q4). The shaded area represents the 0.95 robust confidence intervals for the estimate that uses the instrument that is constructed from forecast errors of all sample countries (black line). VAR specification has 5 lags and endogenous variables have been detrended prior to estimation. BP_g *and* BP_r *refers to Blanchard and Perotti (2002) identification with government spending first (g) and with net revenue first (r) respectively.* CK_g *and* CK_r *refers to Caldara and Kamps (2017) style identification with government spending first (g) and with net revenue first (r) respectively.*

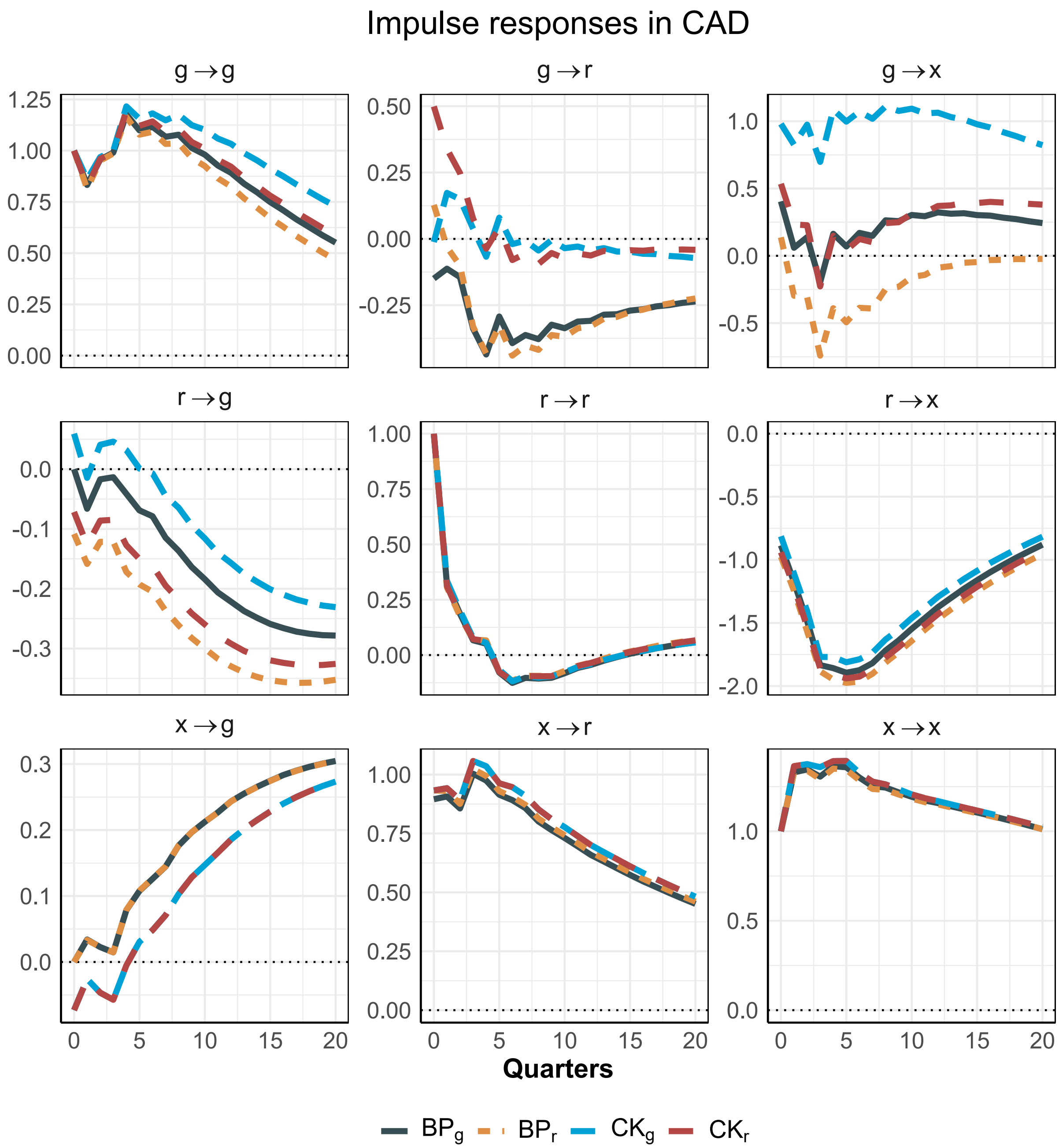

Figure 4: Impulse responses using different identification schemes in the SVAR, Canada.

Notes: Each panel reports impulse responses of a certain endogenous variable from a shock in a certain variable (shock → response). All impulse responses are from the baseline reduced from VAR (5 lags, detrended). However, the different lines in each panel represents the four different identification schemes we use. The impulse responses are scaled and correspond to changes in Canadian dollars. BP_g *and* BP_r *refers to Blanchard and Perotti (2002) identification with government spending first (g) and with net revenue first (r) respectively.* CK_g *and* CK_r *refers to Caldara and Kamps (2017) style identification with government spending first (g) and with net revenue first (r) respectively.*

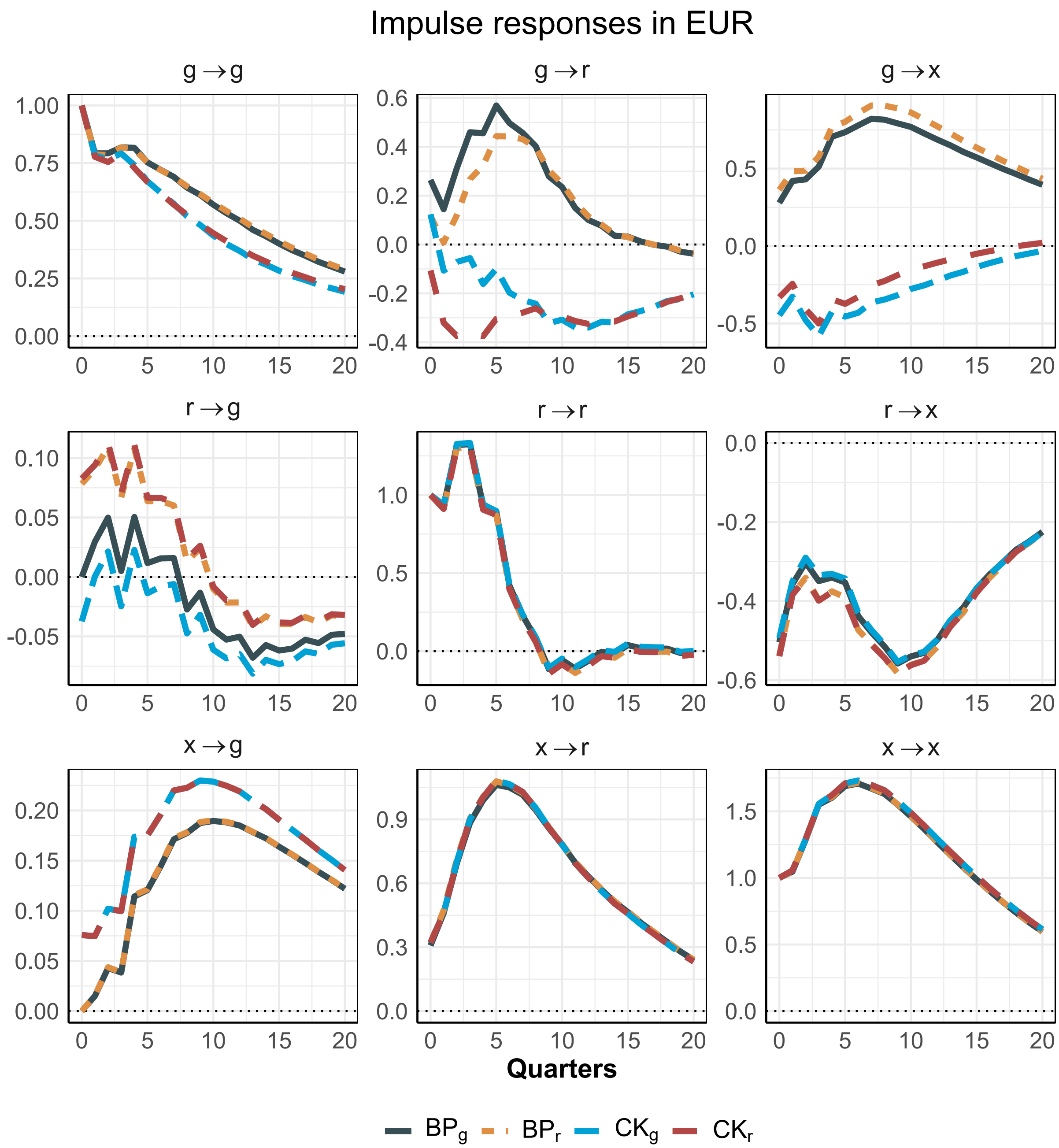

Figure 5: Impulse responses using different identification schemes in the SVAR, Finland.

Notes: Each panel reports impulse responses of a certain endogenous variable from a shock in a certain variable (shock → response). All impulse responses are from the baseline reduced from VAR (5 lags, detrended). However, the different lines in each panel represents the four different identification schemes we use. The impulse responses are scaled and correspond to changes in Euros. BP_g *and* BP_r *refers to Blanchard and Perotti (2002) identification with government spending first (g) and with net revenue first (r) respectively.* CK_g *and* CK_r *refers to Caldara and Kamps (2017) style identification with government spending first (g) and with net revenue first (r) respectively.*

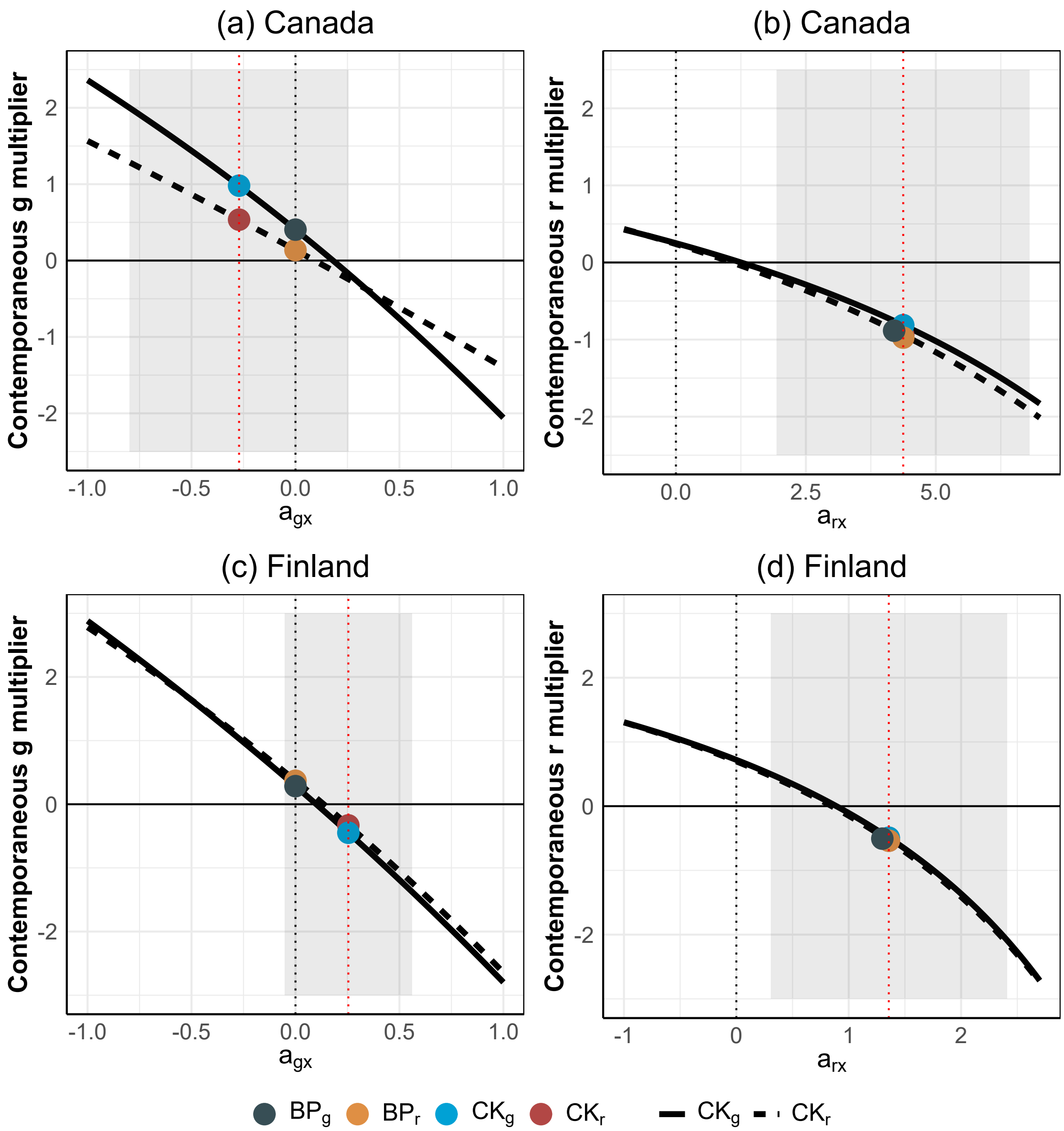

Figure 6: Contemporaneous impulse responses (impact multipliers) of output to fiscal variables as a function of respective output elasticities of fiscal variables

Notes: In each panel the contemporaneous government spending (g) or revenue (r) multiplier is plotted as a function of the 2SLS estimatable parameter a_{gx} *or* a_{rx}*. Solid and dashed lines plot the contemporaneous responses given a parameter value in* CK_g *and* CK_r *identifications while colored dots represent the estimated values across different identification schemes.* BP_g *and* BP_r *identifications set* a_{gx} *to zero (dark dotted line). The red dotted line represents the estimated value of the parameter using either* CK_g *or* CK_r *identification. Shaded area represents 95% confidence interval of the parameter estimate. The VAR specification has 5 lags and detrended endogenous variables. In estimating parameter* a_{gx} *or* a_{rx}*, the output residual is instrumented by trading partner GDP forecast errors.* BP_g *and* BP_r *refers to Blanchard and Perotti (2002) identification with government spending first (g) and with net revenue first (r) respectively.* CK_g *and* CK_r *refers to Caldara and Kamps (2017) style identification with government spending first (g) and with net revenue first (r) respectively.*

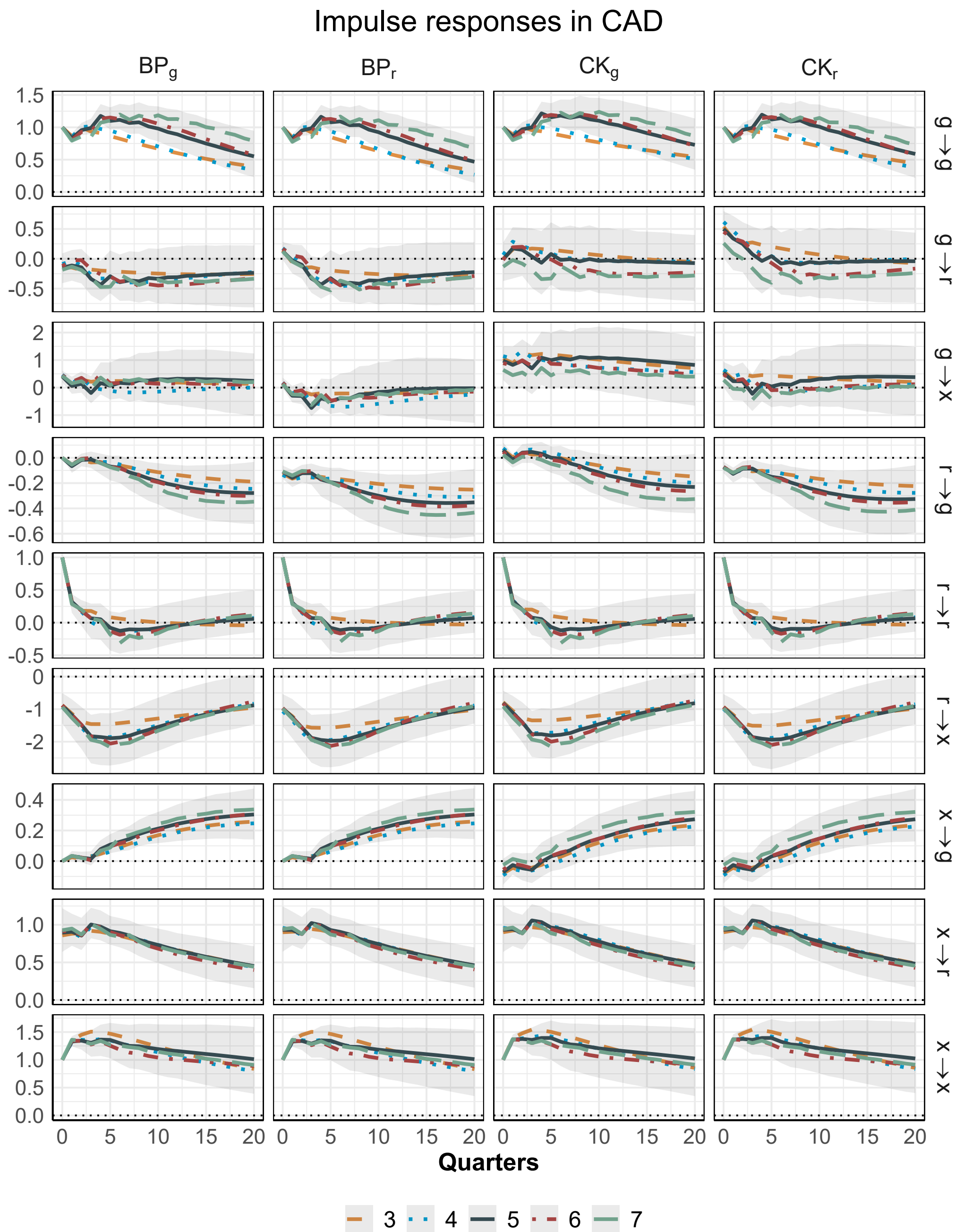

Figure 7: SVAR impulse responses using different lags lengths in the VAR, Canada

Notes: Each column represents one of the four SVAR-IV identification schemes. BP_g *and* BP_r *refers to Blanchard and Perotti (2002) identification with government spending first (g) and with net revenue first (r) respectively.* CK_g *and* CK_r *refers to Caldara and Kamps (2017) style identification with government spending first (g) and with net revenue first (r) respectively. Each panel reports impulse responses of a certain endogenous variable from a shock in a certain variable (shock → response). The different lines in each panel represents impulse responses from a VAR with a different number of lags. In all cases the data is detrended. The grey area represents 68% confidence intervals for the specifications with 5 lags which is constructed by residual-based moving block bootstrap with 2000 draws. The impulse responses are scaled and correspond to changes in Canadian dollars.*

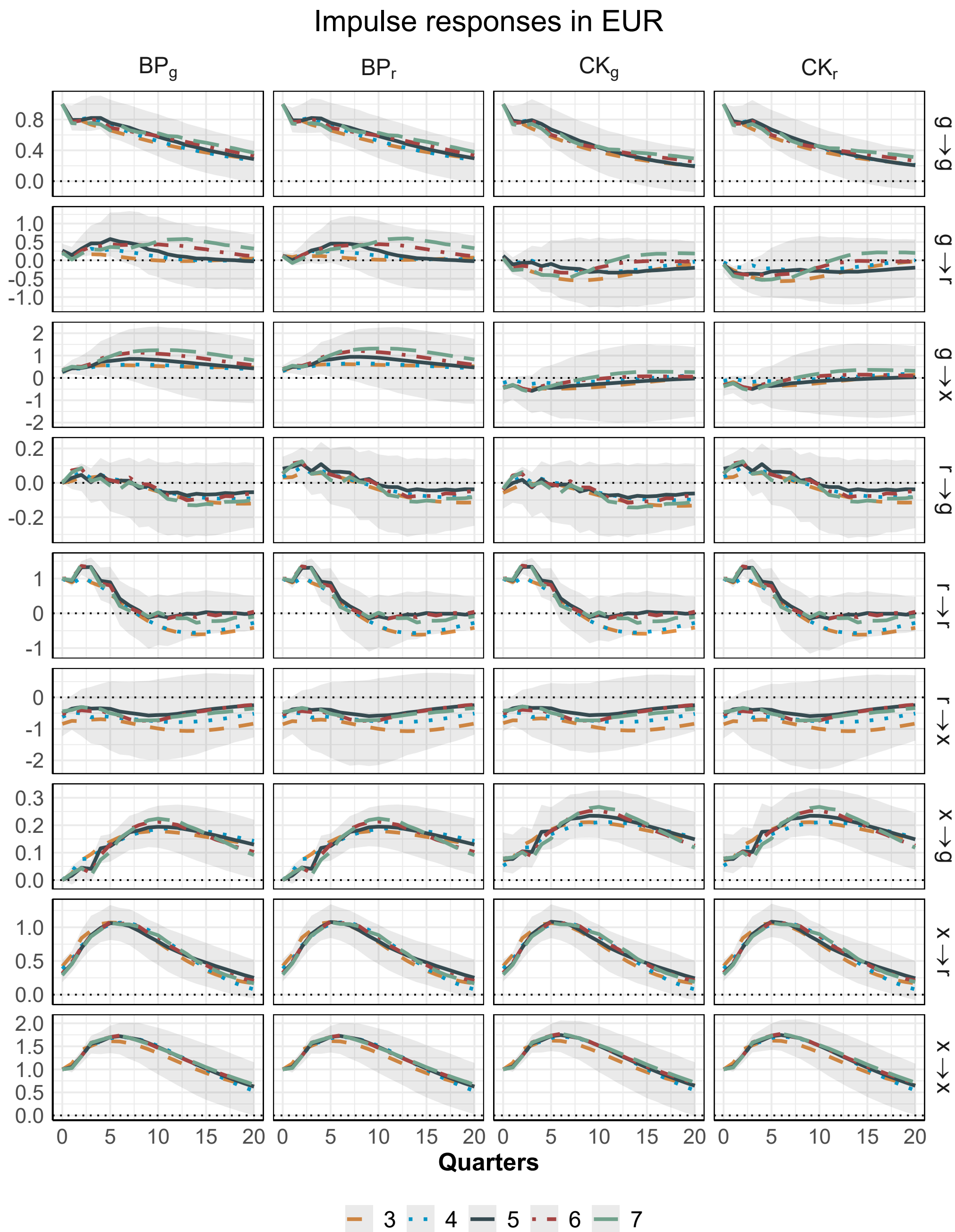

Figure 8: SVAR impulse responses using different lags lengths in the VAR, Finland

Notes: Each column represents one of the 4 different SVAR-IV identification schemes. BP_g and BP_r refers to Blanchard and Perotti (2002) identification with government spending first (g) and with net revenue first (r) respectively. CK_g and CK_r refers to Caldara and Kamps (2017) style identification with government spending first (g) and with net revenue first (r) respectively. Each panel reports impulse responses of a certain endogenous variable from a shock in a certain variable (shock → response). The different lines in each panel represents impulse responses from a VAR with a different number of lags. In all cases the data is detrended. The grey area represents 68% confidence intervals for the specifications with 5 lags which is constructed by residual-based moving block bootstrap with 2000 draws. The impulse responses are scaled and correspond to changes in Euros.

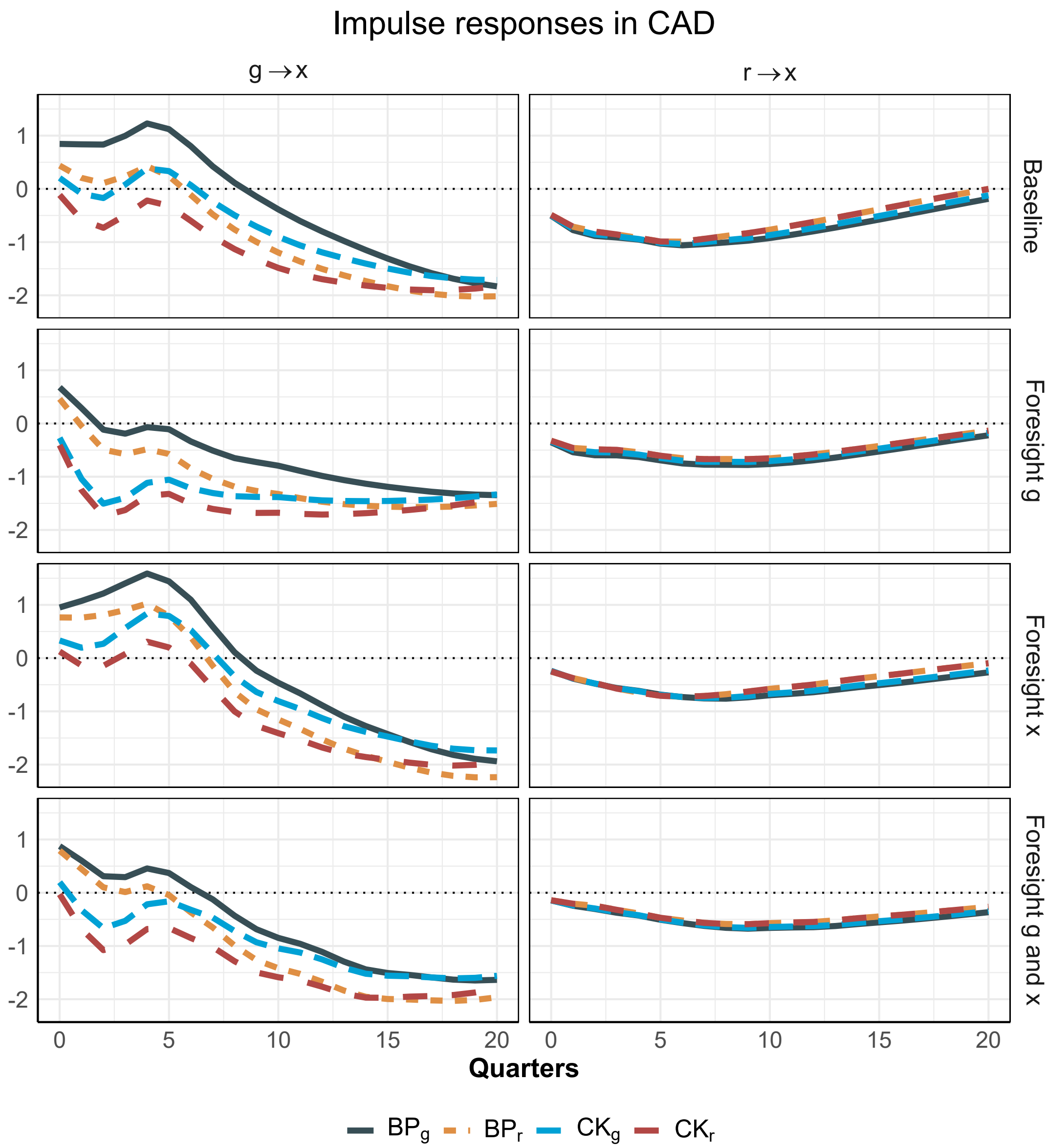

Figure 9: SVAR impulse responses with or without foresight variables, Canada 1987Q1-2013Q4

Notes: In each panel the lines represent the four SVAR-IV identification schemes. BP_g *and* BP_r *refers to Blanchard and Perotti (2002) identification with government spending first (g) and with net revenue first (r) respectively.* CK_g *and* CK_r *refers to Caldara and Kamps (2017) style identification with government spending first (g) and with net revenue first (r) respectively. The right column shows the impulse responses of output to a shock in government spending and left column shows the output responses to a shock in net revenue. Each row contains results from a different VAR specification. The first row shows results from the baseline specification with no foresight variable. The rows 2-4 each show impulse responses from a VAR with foresight variable(s). The VAR in the second row contains a government spending foresight variable whereas the VAR in the third row contains a output foresight variable. The VAR in the last row contains both a government spending and output foresight variables. All specifications contain 5 lags and detrended endogenous variables. The impulse responses are scaled and correspond to changes in Canadian dollars.*

Appendices

Appendix A - Bootstrap confidence intervals

We form confidence intervals by residual-based recursive-design bootstrap (for more see, for example, Kilian and Lüthkepohl 2017). The idea is to draw bootstrap samples from the recursive VAR residuals and then recursively generate bootstrap realizations of the variables in the VAR. To generate these realizations recursively from the estimated VAR one also needs to set some initial values for the VAR variables.

Recently much attention is shown towards the inference of impulse response functions. Jentsch and Lunsford (2019) claim that the confidence intervals shown in Mertens and Ravn (2013) are too small because the wild bootstrap Mertens and Ravn (2013) utilize is asymptotically invalid. Jentsch and Lunsford (2019) propose that a valid residual-based bootstrap method should be preferred against wild bootstrap. Accordingly, such a method is the residual-based moving block bootstrap proposed by Brüggemann, Jentsch and Trenkler (2016). Jentsch and Lunsford (2019) show that the impulse responses in Mertens and Ravn (2013) are not statistically significant under asymptotically valid confidence intervals even within the 68 % significance level. However, they only consider standard Efron's (1979) percentile intervals. That is, the intervals are calculated directly from the sample of bootstrapped impulse responses. Mertens and Ravn (2019) argue in a reply to this critic that the simulation evidence that supports the use of Brüggemann, Jentsch and Trenkler (2016) bootstrap method in finite samples bases on Hall's (1992) percentile intervals which allows for possible finite sample bias. Using Hall's (1992) percentile intervals and residual-based moving block bootstrap Mertens and Ravn (2019) show that their impulse responses are again significantly different from zero at the 68 % level. However, the 95 % significance level that Mertens and Ravn (2013) originally find can not be recovered.

Based on the above discussion we utilize residual-based moving block bootstrap proposed by Brüggemann, Jentsch and Trenkler (2016) and use percentile intervals that are not sensitive for possible finite sample bias. We also amend the block-design to include the instrument series as Jentsch and Lunsford (2019) propose. Contrary to the standard recursive-design bootstrap the block-design allows for heteroskedasticity in the error terms. However, in a SVAR it does not allow for serial correlation.

Kilian and Lüthkepohl (2017) state that, while Efron's (1979) percentile intervals are the most commonly used bootstrap confidence intervals, when modeling structural impulse responses, these should be used with caution. This is because the finite sample distribution of structural impulse response estimators is not necessarily normal. Especially, the possible finite sample bias reflects into the accuracy of Efron's (1979) percentile intervals. Accordingly, this leads to low coverage of the bootstrap confi-dence intervals and furthermore it is not uncommon that the impulse response estimates set outside of the bootstrapped 95 % confidence interval. Kilian and Lüthkepohl (2017) explain that in many applications the assumption, that parameter estimates of a VAR are not systematically different from their true values, is violated because of the finite sample bias. Moreover, the bootstrap estimates tend to amplify the overall bias. This leads to a bootstrap distribution that is centered in a 'wrong place'.

Kilian and Lüthkepohl (2017) review percentile intervals that allow for bias and asymmetry. They state, that in finite samples Hall's (1992) percentile intervals are often less accurate than the percentile t-intervals. Furthermore, Hall's (1992) percentile intervals are not systematically more accurate than Efron's (1979) standard percentile intervals. Therefore, we infer that percentile t-intervals are most accurate confidence intervals in finite samples. The percentile t-intervals Kilian and Lüthkepohl (2017) review, are equal-tailed percentile-t intervals (Efron, 1982) and symmetric percentile-t intervals (Hall, 1992). These are computationally heavy since a nested bootstrap round is needed in order to compute a bootstrap standard error estimate for each individual bootstrap impulse response estimates. According to Kilian and Lüthkepohl (2017), the symmetric percentile-t interval is often but not always more accurate than the equal-tailed percentile-t interval.[27]

Another way to improve confidence interval accuracy which Kilian (1998) proposes is to first correct for finite sample bias in the least squares VAR parameter estimates by bootstrap and then form bootstrap intervals using Efron's (1979) percentiles (for more see Kilian and Lüthkepohl (2017)). Kilian and Lüthkepohl (2017) note that in typical sized samples bias adjustment is not necessarily enough to ensure accurate inference when the VAR model includes a deterministic trend. The simulation results in Kilian (1999) recommend the use of Kilian's (1998) method in finite samples.

After all, it is not completely clear which method to use over another in a finite sample while the (stylised) simulation based evidence supports the use of Kilian (1998). Given this and the above discussion we utilize the beforehand bias correction proposed by Kilian (1998) with Efron's (1979) percentiles.

Appendix B - Extended table on relevancy and exogeneity

[27] The review concerning the different percentile intervals given in Kilian and Lüthkepohl (2017) raises an anxiety that Jentsch and Lunsford (2019) and Mertens and Ravn (2019) both still possibly draw their confidence intervals inaccurately. Moreover, Brüggemann, Jentsch and Trenkler (2016) test their bootstrap method only using Hall's (1992) percentile intervals. It would be very useful to see how this method behaves in finite samples, for example, when percentile-t intervals are used.

Panel A: Relevance

	Dependent variable: Forecast error of output							*Dependent variable:* Contemporaneous output						
	CAN (1)	CAN (2)	CAN (3)	FIN (4)	FIN (5)	USA (CK) (6)	USA (7)	CAN (8)	CAN (9)	CAN (10)	FIN (11)	FIN (12)	USA (CK) (13)	USA (14)
GDP (SPF)	0.32***							0.33***						
	(0.08)							(0.08)						
GDP (OECD)		0.37***		1.66***					0.35***		1.16***			
		(0.09)		(0.56)					(0.10)		(0.29)			
Imports (OECD)			0.11***		0.49***					0.13***		0.25**		
			(0.02)		(0.17)					(0.03)		(0.11)		
Util. adj. TFP						0.43***	0.28***						0.54***	0.22***
						(0.10)	(0.07)						(0.07)	(0.06)
N	200	176	176	88	88	152	188	200	176	176	171	171	224	275
Adjusted R^2	0.23	0.26	0.28	0.44	0.40	0.15	0.11	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Panel B: Exogeneity

	CAN			FIN		USA		CAN			FIN		USA	
	GDP (SPF) (1)	GDP (OECD) (2)	Imports (OECD) (3)	GDP (OECD) (4)	Imports (OECD) (5)	Util. adj. TFP (CK) (6)	Util. adj. TFP (7)	GDP (SPF) (8)	GDP (OECD) (9)	Imports (OECD) (10)	GDP (OECD) (11)	Imports (OECD) (12)	Util. adj. TFP (CK) (13)	Util. adj. TFP (14)
FE of g (OECD)	0.01	−0.03	−0.18	0.04	0.01									
	(0.07)	(0.06)	(0.20)	(0.03)	(0.11)									
FE of g (SPF)						−0.02	0.28***							
						(0.10)	(0.09)							
g								−0.08	−0.01	−0.22	0.04*	0.11	−0.03	0.13***
								(0.06)	(0.06)	(0.24)	(0.02)	(0.07)	(0.05)	(0.05)
N	132	132	132	88	88	101	137	200	176	176	171	171	224	275
Adjusted R^2	0.14	0.12	0.21	0.23	0.36	0.16	0.12	0.06	0.10	0.17	0.14	0.13	0.13	0.11

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table 6: Relevance and exogeneity.

Notes: In Panel A columns 1-7 report OLS estimation results from a regression of professional forecast errors of the quarterly growth of GDP on different instruments. Columns 1-7 are the same as show in Table 2. Columns 8-14 provide additional results from a regression which tests the correlation between our instrument and the reduced from baseline VAR output residuals. For Canada and Finland forecast errors of GDP are from OECD Economic Outlooks and for the US they are from the Survey of Professional forecasters (SPF). Instrument GDP (SPF) is constructed from US GDP forecast errors (SPF) and GDP (OECD) and Import (OECD) instruments from a export-share weighted mean of OECD trading partner forecast errors (OECD). Utilization-adjusted TFP series of Fernald (2014) is either from Caldara and Kamps or from the November 14th 2019 vintage and downloaded from `https://sites.google.com/site/fernaldjg/TFP`*. In Panel B, columns 1-7 report OLS estimation results from a regression of different instrument (now the dependent variable) on the professional forecast errors in the growth of the sum of general government consumption and investment. Columns 1-7 are the same as show in Table 2. Columns 8-14 provide additional results from a regression where we regress government spending against our instrument and use lags of the endogenous variables as controls. If we assume, as Blandchard and Perotti (2002), that government spending has no contemporaneous relation with unexpected output and revenue shocks this regression tests the correlation between our instrument and the structural government spending shocks. For Canada and Finland forecast errors of government spending are from the OECD Economic Outlooks and for the US they are from the Survey of Professional forecasters (SPF). In both panels, each model includes lags of VAR endogenous variables as controls; 5 for Canada and Finland, 4 for the USA. Robust Eicker–Huber–White standard errors are used throughout.*

Appendix C - Additional results on the use of Fernald's (2014) utilization-adjusted TFP series as an instrument

In this appendix we study the different vintages of the Fernald (2014) TFP series. We try to test the relevancy and exogeneity of this series if it is concidered as an TFP instrument in a similar fashion as in the main text in section 4. Note that the newest vintages of this series should be the most accurate version of it.

In Figure 10 we plot coefficient estimates from similar type of regressions as in Tables 2 for different vintage of the Fernald (2014) TFP series. In panel A we report the coefficients and their standard error bands for relevance and in panel B the coefficients and their standard error bands for exogeneity.

In panel A the coefficients before 2014 vintages are all clearly significantly different from zero. After that while significant there is a clear decrease in the size of the coefficient towards zero. Similarly in panel B before 2014 the coefficients are sometimes insignificant. But after 2014 the coefficients are significant indicating possible violation of the exogeneity assumption. Note that we mark the Caldara and Kamps (2017) instrument with yellow in Figure 10.

Furthermore, in Figure 12 we show the weak instrument test (see section 4) results for the TFP instrument. For these calculations we use the Caldara and Kamps (2017) data. The earlier vintages of the instrument are not weak whereas sometime during 2015 the revisions have weakened the instrument. Yet, in Figure 11 we depict estimated values of certain structural parameter of the identification schemes used in this study (see section 4 and 5). Again depending on the vintage one gets very different sized coefficient estimates for these parameters. Furthermore, none of them are quite similar with the ones we get with the Caldara and Kamps (2017) instrument.

One more interesting and puzzling thing is that Caldara and Kamps (2017) refer to Fernald (2014) when describing the forming of their TFP instrument. However, we are not able to find practically any correlation between any revision of Fernald's (2014) TFP series and the instrument Caldara and Kamps (2017) use. Caldara and Kamps (2017) do not provide a detailed description how they obtain their instrument. Anyway, it is not directly one of the series Fernald (2014) provides.[28] Assuming that Caldara and Kamps (2017) only used the same method as Fernald (2014) and not directly one of the ready-made series one would still expect strong correlation between these series. Furthermore, those series that Fernald (2014) provides are related with the forecast errors which suggest that these series

[28]One possibility is that Caldara and Kamps (2017) use some old revision or earlier version of the Fernald's (2014) TFP series which is not available presently.

as instruments might not be exogenous. Contrary, the instrument Caldara and Kamps (2017) use anyway seems to be exogenous according to this test, despite that it should be a Fernald (2014) style TFP series.

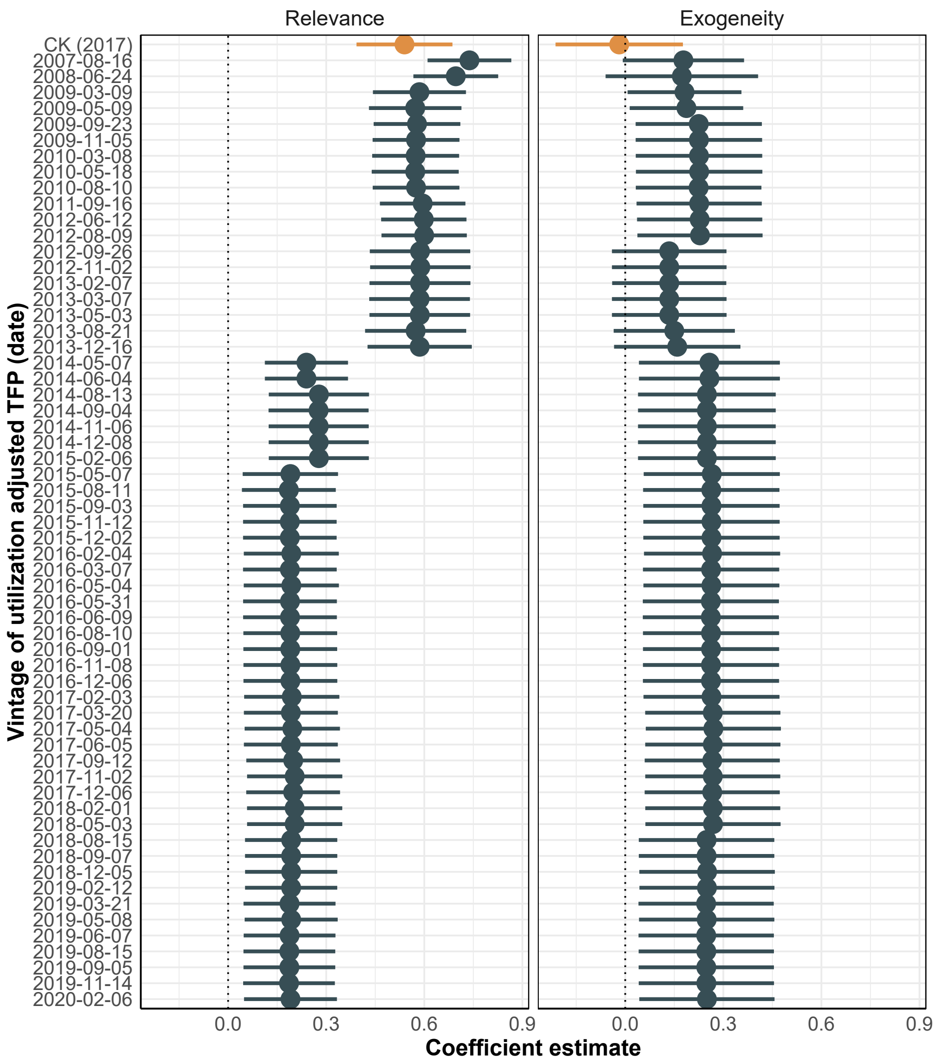

Figure 10: Coefficient estimates from repeated regressions of Table 2 on the relevance and exogeneity of the utilization-adjusted TFP instrument for output using US data.

Notes: The first column plots the point estimates from repeated regressions of real GDP on the utilization adjusted TFP using data from different vintages of the series of Fernald (2014) except in the first row where data from Caldara and Kamps (2017) is used. The second column plots estimates from a model where the utilization adjusted TFP series is regressed on forecast errors of government spending. Models in both columns include four lags of real GDP, government spending and revenues as controls. Errorbars plot the 95% confidence intervals of the estimated coefficients. Forecast errors of government spending are constructed from the Survey of Professional Forecasters data. Robust Eicker-Huber-White standard errors are used throughout.

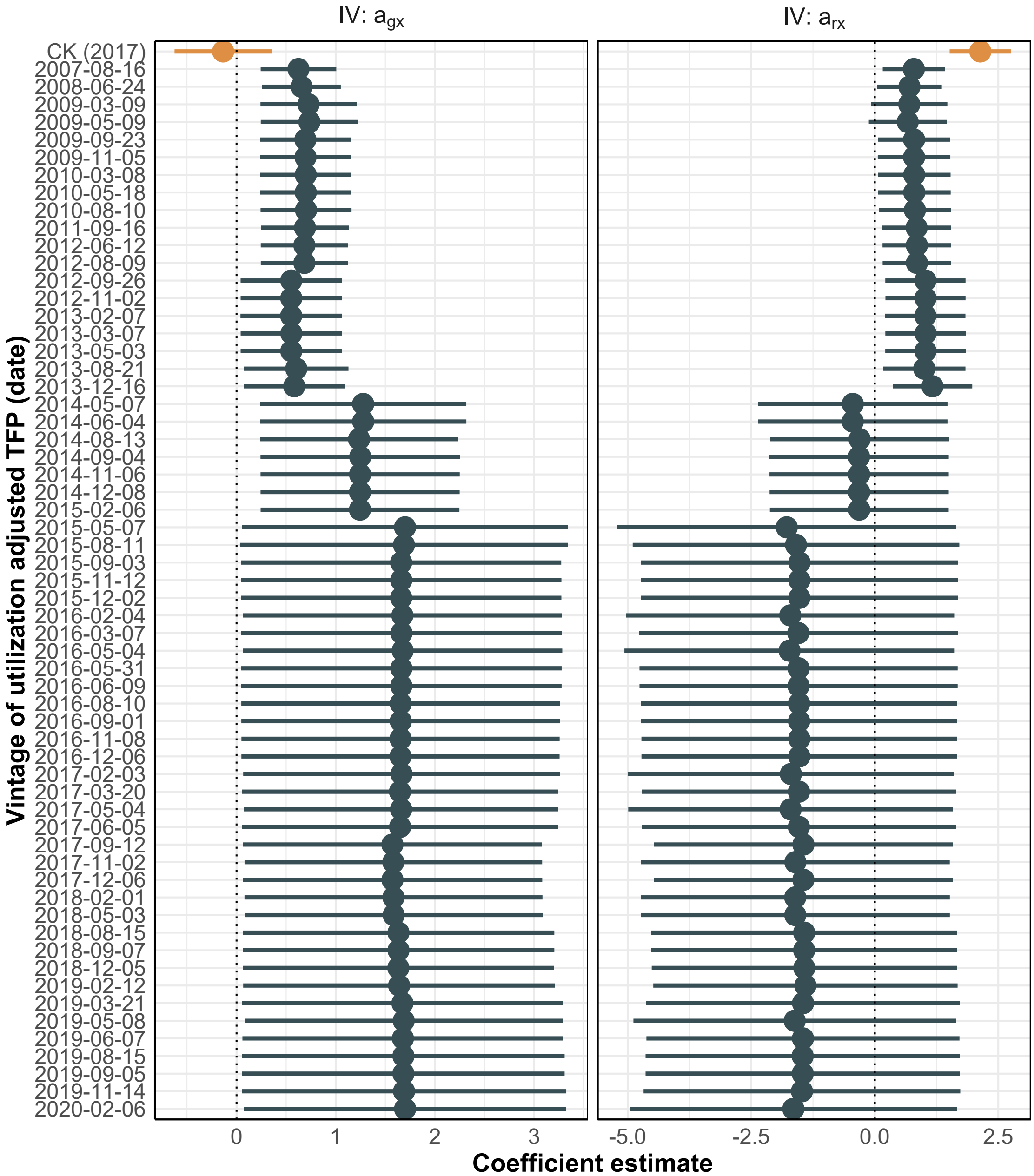

Figure 11: Estimates of certain structural parameters from SVAR-IV using US data and different vintages of Fernald's (2014) TFP series as an instrument.

Notes: The columns plot the point estimates of certain structural parameters from a SVAR-IV where the adjusted TFP instrument is utilized. Estimates for the output elasticity of net revenue are shown in the right column and estimates for the output elasticity of government spending are shown in the left column. The different estimates represent different vintages of the TFP series of Fernald (2014) except in the first row where data from Caldara and Kamps (2017) is used. Models in both columns include four lags of real GDP, government spending and revenues as controls. Errorbars plot the 95% confidence intervals of the estimated coefficients. Forecast errors of government spending are constructed from the Survey of Professional Forecasters data. Robust Eicker-Huber-White standard errors are used throughout.

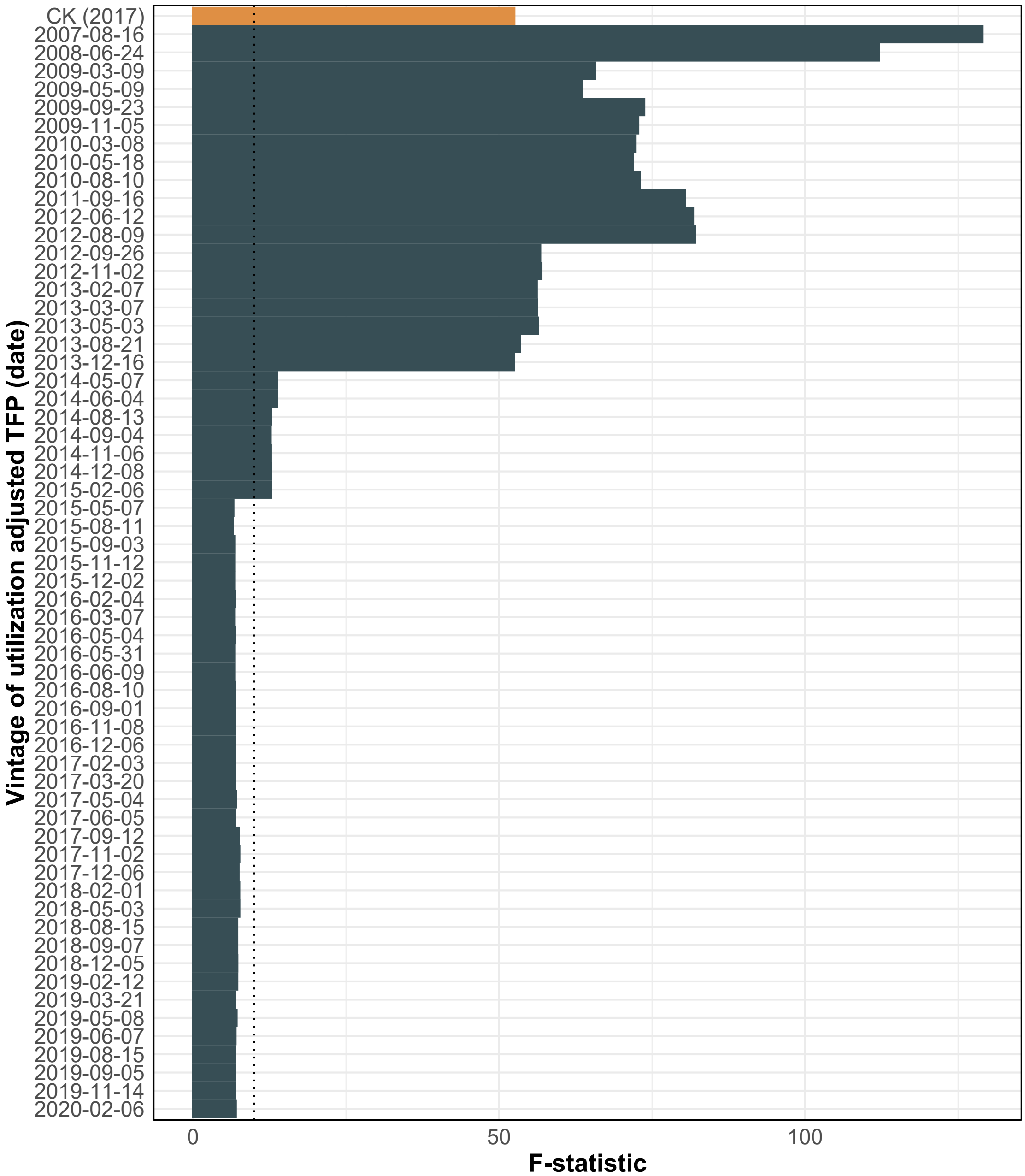

Figure 12: The first stage F-statistics for the output elasticity of net revenue estimates of the SVAR-IV using US data and different vintages of Fernald's (2014) TFP series as an instrument.

Notes: Each bar represents the first stage efficient F-statistics related with the output elasticity of net revenue estimates from SVAR-IV using US data and different vintages of Fernald's (2014) TFP series as an instrument. The fisrt bar (orange) represents a model where data from Caldara and Kamps (2017) is used.

Appendix D - Structural identification matrices

The equations presented below between the reduced form VAR residuals u_j and structural shocks e_j ($j \in \{g, r, x\}$) correspond to the four different identification schemes used across the paper. The equations are represented here in the conventional AB-matrix form with the LHS A-matrix containing the contemporaneous elasticities between model variables and RHS B-matrix containing the direct effects of structural shocks to model variables.

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -a_{rx} \\ -a_{xg} & -a_{xr} & 1 \end{bmatrix} \begin{bmatrix} u^g \\ u^r \\ u^x \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ b_{rg} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} e^g \\ e^r \\ e^x \end{bmatrix} \tag{BP$_g$}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -a_{rx} \\ -a_{xg} & -a_{xr} & 1 \end{bmatrix} \begin{bmatrix} u^g \\ u^r \\ u^x \end{bmatrix} = \begin{bmatrix} 1 & b_{gr} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} e^g \\ e^r \\ e^x \end{bmatrix} \tag{BP$_r$}$$

$$\begin{bmatrix} 1 & 0 & -a_{gx} \\ 0 & 1 & -a_{rx} \\ -a_{xg} & -a_{xr} & 1 \end{bmatrix} \begin{bmatrix} u^g \\ u^r \\ u^x \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ b_{rg} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} e^g \\ e^r \\ e^x \end{bmatrix} \tag{CK$_g$}$$

$$\begin{bmatrix} 1 & 0 & -a_{gx} \\ 0 & 1 & -a_{rx} \\ -a_{xg} & -a_{xr} & 1 \end{bmatrix} \begin{bmatrix} u^g \\ u^r \\ u^x \end{bmatrix} = \begin{bmatrix} 1 & b_{gr} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} e^g \\ e^r \\ e^x \end{bmatrix} \tag{CK$_r$}$$

Appendix E - Estimates for all structural parameters

In Tables 3 and 4 we present regression results for all of the key structural parameters across each of the four different identification schemes. Each of the columns represents an individual regression that belongs to one of the four identification schemes. The columns have been divided under each identification scheme and are ordered such that for each type of identification the sequential order of regressions runs from left to right. The regressors are placed row-wise in the table. If a row contains a value at certain column this means that this regressor is included in the regression which the column represents. All structural parameters in the table are retrieved from regressions where we control for 5 lags of the detrended VAR endogenous variables in accordance with the baseline reduced form VAR specification. We also depict these estimated parameters in Figures 13 and 14. In these figures the dots represent the estimates and the lines their 0.95 confidence intervals. The different identification schemes are marked with different colors.

The estimates of the output elasticity of net revenues that are gathered on row x in Tables 3 and 4 are similar across all four identification schemes. Accordingly, also the estimate of the contemporaneous effect of revenues on output is very similar in each identification scheme. Indeed the differences in results between BP style identifications and CK style identifications springs from the parameters that relate with government spending. The parameter gathering the contemporaneous response of government spending to output is restricted to zero in BP identification while in CK identification this estimate is allowed to be non-zero. While the 2SLS estimates for this parameter are not statistically significantly different from zero at the 0.05 level, allowing it to be non-zero affects the later stages of estimation where the resulting identified government spending shock series is used. In relation to this, the other key parameter which differs between BP and CK is the effect of government spending on output. For Canada this estimate is positive and significant in CK identification and insignificant in BP identification whereas for Finland this estimate is positive and significant (at 0.1 level) in BP identification and insignificant and negative in CK identification.

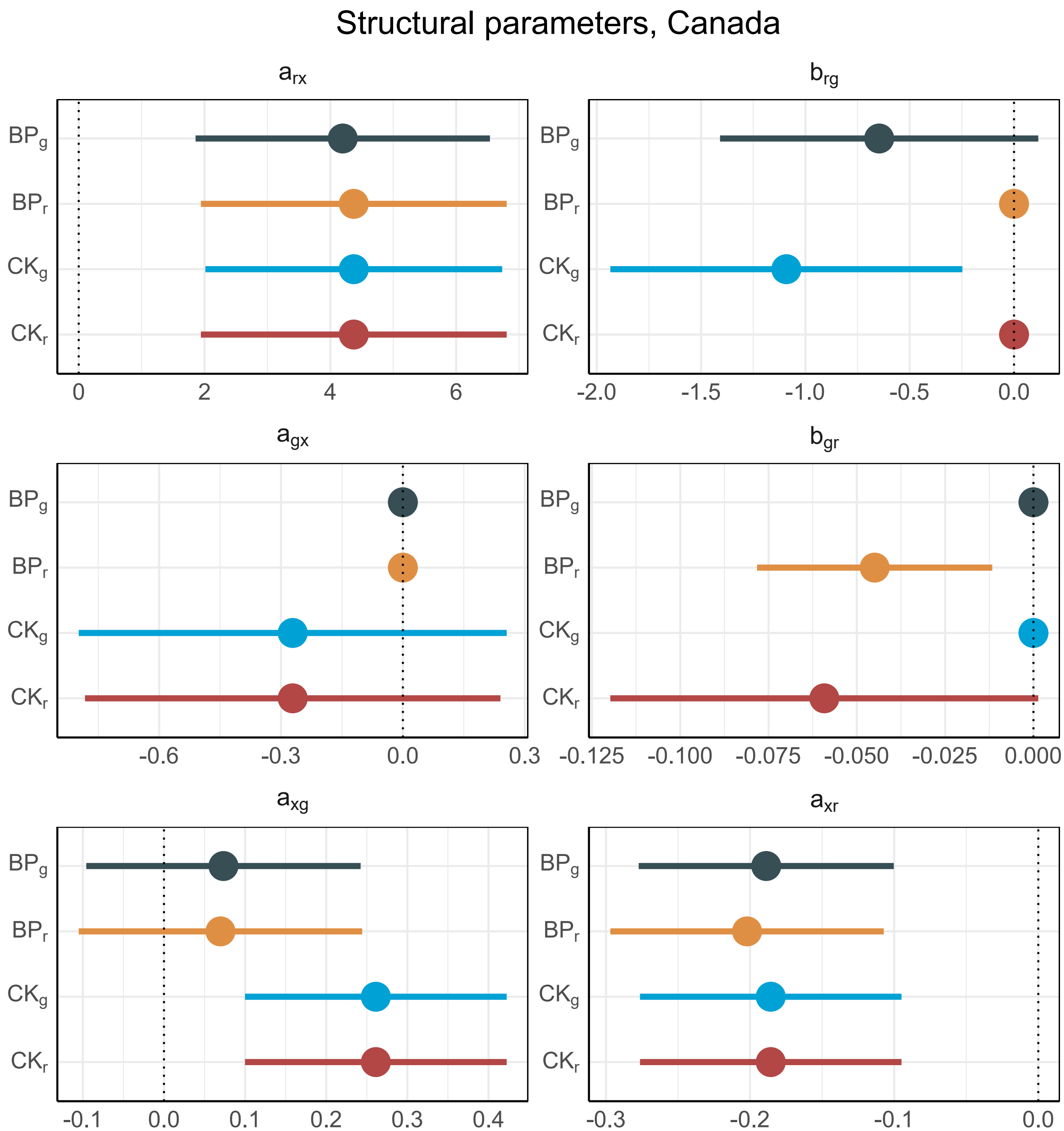

Figure 13: Structural parameters across different identifications, Canada

Notes: Each panel reports the structural parameter estimates and their 95 % confidence bands from the four identification schemes we utilise. The VAR specification has 5 lags and endogenous variables have been detrended prior to estimation. Output is instrumented with US (SPF) forecast error of GDP. BP_g *and* BP_r *refers to Blanchard and Perotti (2002) identification with government spending first (g) and with net revenue first (r) respectively.* CK_g *and* CK_r *refers to Caldara and Kamps (2017) style identification with government spending first (g) and with net revenue first (r) respectively.*

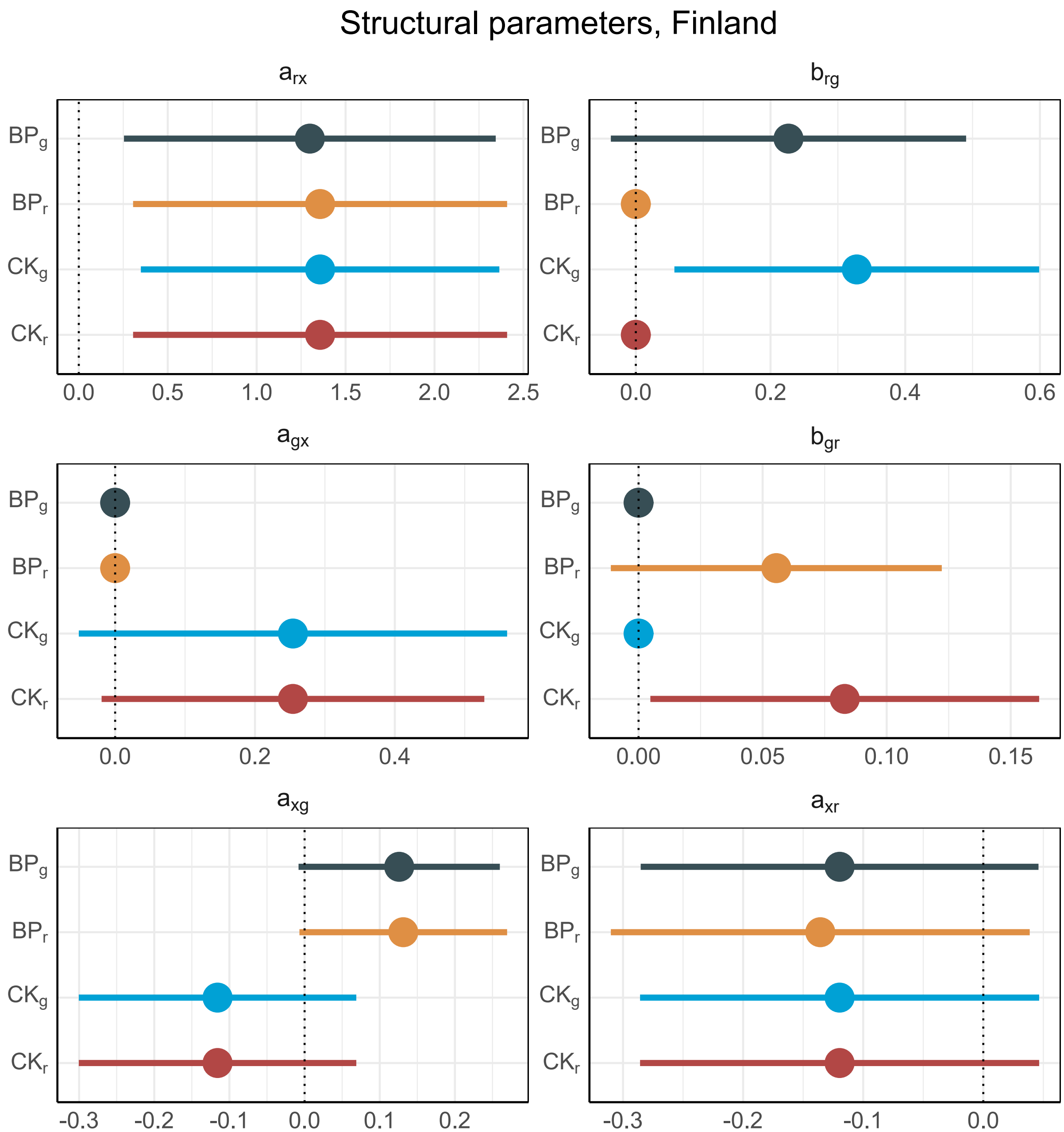

Figure 14: Structural parameters across different identifications, Finland

Notes: Each panel reports the structural parameter estimates and their 95 % confidence bands from the four identification schemes we utilise. The VAR specification has 5 lags and endogenous variables have been detrended prior to estimation. Output is instrumented with weighted trading partner (OECD) forecast error of GDP. BP_g *and* BP_r *refers to Blanchard and Perotti (2002) identification with government spending first (g) and with net revenue first (r) respectively.* CK_g *and* CK_r *refers to Caldara and Kamps (2017) style identification with government spending first (g) and with net revenue first (r) respectively.*

Appendix F - Robustness to reduced form VAR specification

In this appendix we show that the elasticity estimates are robust to different choices related with the specification of the reduced form VAR. As before, we include 5 lags in our baseline specification of the reduced form VAR. In Figures 7 and 8 we show how the elasticity estimate changes if different VAR specifications are considered. The two different columns report results from VAR with detrended endogenous variables as in the baseline specification and alternatively with endogenous variables in log levels. The first row shows the coefficient estimates and the corresponding 0.95 robust confidence intervals. The second row shows the first stage robust F-statistics from each corresponding regression. Additionally, the third row shows the associated Wu-Hausman statistic's p-value.[29] In Figures 7 and 8 the different colors represent the four different identifications considered. In each specification we utilize the GDP based forecast error instrument. For Canada this is the instrument constructed from the SPF forecast errors and for Finland this is the OECD instrument where all available countries are utilized and weighted according to their share in Finnish exports.

Both Figures 7 and 8 suggest that the specification of the reduced form VAR has only a little effect on the estimate of the elasticity coefficient. Also the associated robust F-statistic, while it varies moderately, is above 10 in each of the specifications. In the case of Finland, as more lags are included in the reduced form VAR, the null of Wu-Hausman test cannot be rejected at the typical 0.05 level. This could simply be because of the moderate sample size as one can still reject the null in most cases at the 0.15 level. For Canada this test's null hypotheses is in all specification clearly rejected.

Overall it seems that the instrument is robust to choices related with VAR specification, such as whether the endogenous variables are in log levels or detrended and also what is the lag order of the reduced form VAR.

[29]Wu-Hausman test's null hypothesis states that both IV and OLS estimates are consistent but the latter is more efficient. If the null is rejected then only IV-estimate is considered to be consistent.

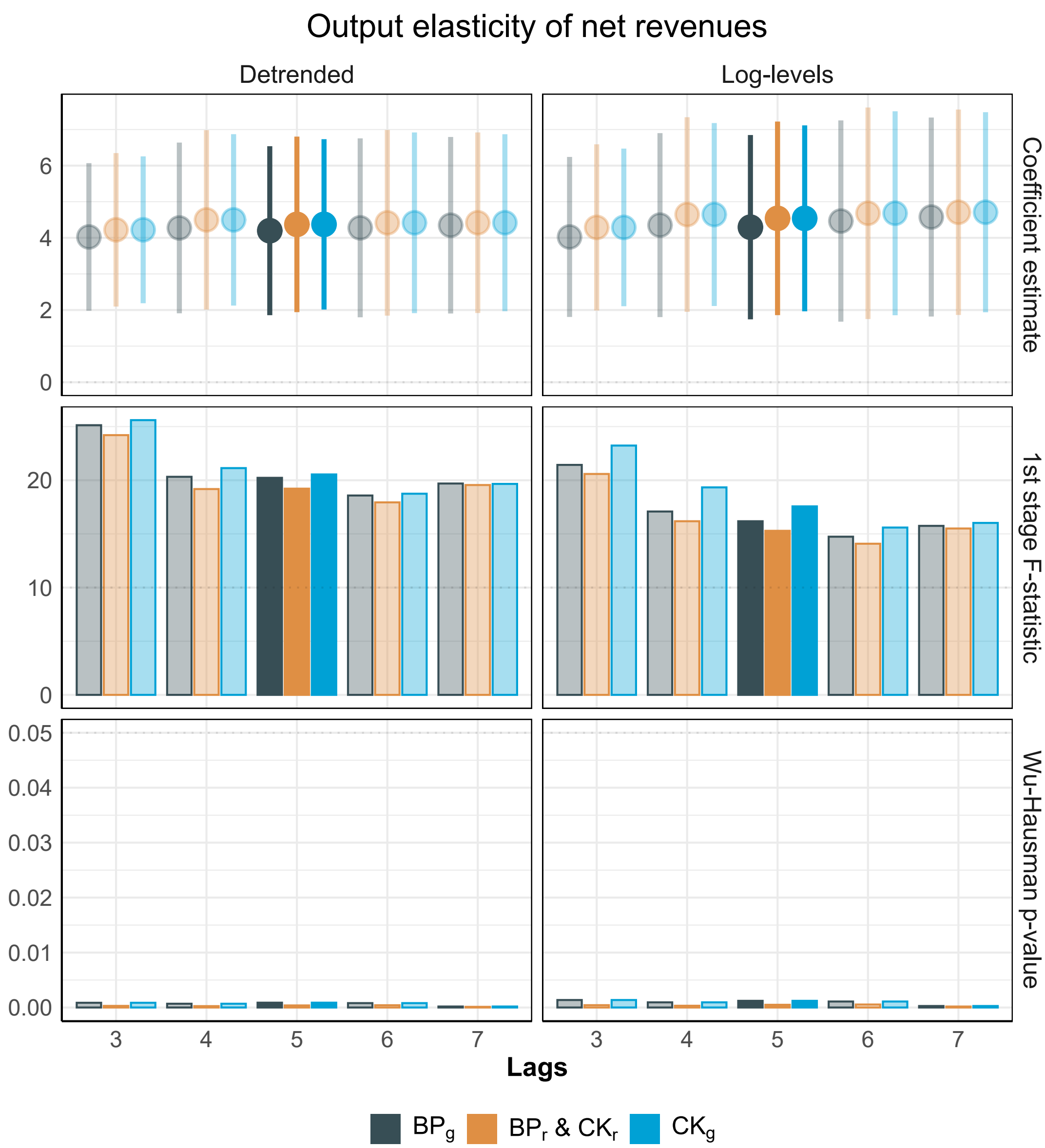

Figure 15: 2SLS estimates of the output elasticity of net revenues, robust 1st stage statistics and Wu-Hausman test p-values using different reduced from VAR specifications, Canada

Notes: The panels in the left column represent VAR specifications with detrended data whereas the results in the right panel represent specifications with data in log-levels. In all panels the different colors represent the different SVAR-IV identification schemes we utilize. Each number on the x-axis represents a VAR specification with the corresponding lag length. The unshaded colors represents our baseline specification (5 lags). In the upper panels we show the coefficient estimates of the output elasticity of net revenues and their 95 % confidence bands we obtain with our instrument. The panels in the middle report the associated first stage efficient F-statistics. The lowest panels report the p-values of the Wu-Hausman test from the first stage regressions. BP_g and BP_r refers to Blanchard and Perotti (2002) identification with government spending first (g) and with net revenue first (r) respectively. CK_g and CK_r refers to Caldara and Kamps (2017) style identification with government spending first (g) and with net revenue first (r) respectively.

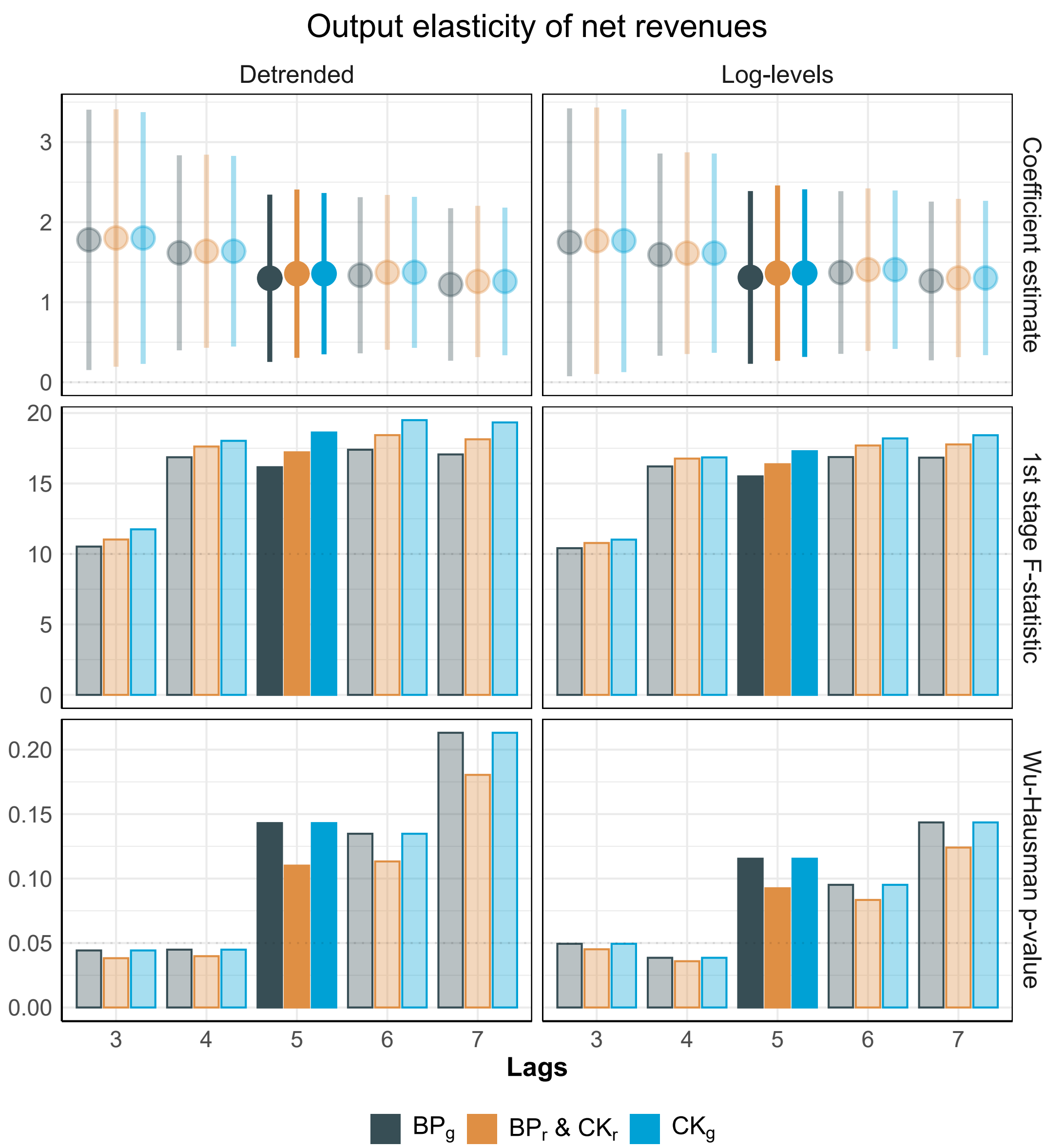

Figure 16: 2SLS estimates of the output elasticity of net revenues, robust 1st stage statistics and Wu-Hausman test p-values using different reduced from VAR specifications, Finland

Notes: The panels in the left column represent VAR specifications with detrended data whereas the results in the right panel represent specifications with data in log-levels. In all panels the different colors represent the different SVAR-IV identification schemes we utilize. Each number on the x-axis represents a VAR specification with the corresponding lag length. The unshaded colors represents our baseline specification (5 lags). In the upper panels we show the coefficient estimates of the output elasticity of net revenues and their 95 % confidence bands we obtain with our instrument. The panels in the middle report the associated first stage efficient F-statistics. The lowest panels report the p-values of the Wu-Hausman test from the first stage regressions. BP_g *and* BP_r *refers to Blanchard and Perotti (2002) identification with government spending first (g) and with net revenue first (r) respectively.* CK_g *and* CK_r *refers to Caldara and Kamps (2017) style identification with government spending first (g) and with net revenue first (r) respectively.*

Appendix G - Fiscal multipliers

In her review of the recent fiscal multiplier literature Ramey (2019) notes that before the financial crisis it was typical to report peak multipliers in the fashion of Blanchard and Perotti (2002). These multipliers represent the ratio of the output response to the initial fiscal impulse at a horizon where this measure achieves its highest value. Ramey (2019) calls these figures quasi multipliers as they do not, as such, represent the dynamic multiplier which would also account for the development of the fiscal variable after the initial response. She points out that the dynamic multiplier is more useful when considering the actual policy making as it directly reflects the total costs and benefits of a fiscal policy plan.

After the financial crisis it has become customary to report multipliers in the fashion of Mountford and Uhlig (2009) who calculate present value cumulative fiscal multipliers. This type of a multiplier represents the dynamic multiplier as it takes into account both the dynamic response of output and the dynamic response of the fiscal variable after an shock to the latter. The multiplier is calculated as the discounted value of the cumulative output response divided by the discounted value of the cumulative government spending response over a certain time horizon. The interest rate used for present value discounting should reflect the risk free rate but in practice also a zero discount rate is often adopted as the horizon of interest is typically quite short and different discount rates have a negligible effect on the final multipliers when reasonable values are used.

Calculation of both peak and dynamic multipliers also includes a scaling factor when log-values of the endogenous variables are used in the VAR model. When log-values are used, the estimated structural parameters can be interpreted as elasticities. Standard practice is then to multiply the impulse responses from a model in logs by a conversion factor in order to achieve impulse responses that correspond to changes in dollars or other monetary units. Typically the conversion factor that is used is based on the sample average ratio of the fiscal variable to output.[30]

Accordingly, we derive the fiscal multipliers reported in this paper following Mountford and Uhlig (2009). For example, the government spending multiplier at horizon h is calculated as

$$Multiplier_h = \frac{\sum_{j=0}^{h} \frac{x_j}{(1+i)^j}}{\sum_{j=0}^{h} \frac{g_j}{(1+i)^j}} \frac{1}{\overline{g/x}}, \tag{7}$$

[30] Ramey (2019) criticises the custom to use the average of GDP divided by government spending as this scaling factor. The average value is problematic as GDP is cyclical but government spending is not. This, for example, can exaggerate the multipliers in recessions. Ramey (2019) mentions two approaches to overcome this problem; either to divide both GDP and government spending by lagged GDP or by a measure of potential GDP.

where h is the time horizon, i is the risk-free rate, x_j is the response of output at period j and g_j is the response of government spending at period j. The scaling factor $\overline{g/x}$ is the sample average of government spending divided by GDP. For simplicity, we use a zero discount rate. As Ramey (2019) mentions, the interest rates typically used produce very similar multipliers.

In this appendix we next report estimates of the cumulative fiscal multiplier across the four different identification schemes for both Canada and Finland. In the case of Canada we also report fiscal multipliers derived from a SVAR model with fiscal foresight variables for both government spending and GDP. The multipliers are calculated as we outline above. For both countries and all specifications we report the instant impact, 1 year and 2 year horizon cumulative multipliers. In each panel we also report the value of the impulse response of government spending and GDP at the point of the horizon of the cumulative multiplier. In parenthesis we report the bootstrapped 68 percentage confidence intervals which we use for inference. Since the four identification schemes produce quite different multiplier estimates, we also discuss the resulting average multipliers calculated from the four estimates in the text. For both countries we consider two different estimation periods. Overall, altering the estimation period has a remarkable effect on the impulse responses and so also the fiscal multipliers.[31]

Estimated fiscal multipliers for Canada are reported in Table 7. The first panel reports estimates using the full-sample and the second panel reports multipliers from a sample with years 1987-2013 included. The third panel reports multiplier also from a sample with years 1987-2013 included but from a SVAR model that includes fiscal foresight variables. The samples in Panels B and C are chosen to cover the years 1987-2013 because this is the period we have data for the fiscal foresight variables. The samples in Panels B and C allow to study the possible effect of the fiscal foresight variables on the multiplier estimates.

In Panel A, the point estimates of the fiscal multiplier using different identification schemes differ from each other ranging from -0.32 to 0.9 with the one year horizon and from -0.34 to 0.9 with the two year horizon. The average value of the multiplier over the four identification schemes is 0.22 with one year horizon and 0.21 with two year horizon. In Panel B the point estimates of the multipliers range from -0.44 to 1 with the one year horizon and from -0.6 to 0.85 with the two year horizon, with average value of 0.24 and 0.07 respectively. In Panel C the point estimates of the multipliers range from -0.83 to 0.65 with the one year horizon and from -0.94 to 0.33 with the two year horizon, with average value

[31]Note that, altering the estimation period does not remove the differences between the different identification schemes. Therefore, the differences between the different identification schemes are not a result of a certain sample.

of -0.05 and -0.28 respectively.

The estimates for Canada vary, and there seems to be no clear patterns to be found. For example, it seems that the multipliers from BP and CK identification do not differ systematically in size in the case of Canada. Note that for the full sample the multipliers from CK identification are larger than from BP identification whereas for the shorter sample the multipliers are larger for BP than for CK identification. Overall, the multiplier estimates for Canada are rarely significantly different from zero even at the 0.68 level. This could reflect that actually the fiscal multiplier for Canada might be something close to zero. If one considers for example the canonical Mundell-Fleming model, this is how one would predict a small open economy with a floating exchange rate to react to fiscal policy.[32]

The fiscal multipliers for Finland are reported in Table 8. Panel A reports multipliers from the full-sample and Panel B reports multipliers from a sample with years 1995-2018 included. This division is motivated by the fact that post 1995 period both excludes the deep recession of early 1990s and also better represents the current macroeconomic policy regime in Finland. This is as the Finnish monetary policy and exchange rate regime changed once Finland became a member of the Eurozone in the beginning of 1999.[33]

In Panel A of Table 8 the point estimate of the fiscal multipliers from different identification schemes differ from each other ranging from -0.55 to 0.65 with the one year horizon and from -0.59 to 0.9 with the two year horizon. The average value of the multiplier over the four identification schemes is 0.06 with one year horizon and 0.17 with two year horizon. In Panel B the point estimates of the multipliers range from 0.38 to 2.03 for the one year horizon and from 2.5 to 4.96 for the two year horizon, with average values of 1.22 and 3.75 respectively.

In the case of Finland there seems to be a clear difference in the size of the estimated government spending multipliers between the two identification schemes. The multipliers using CK identification are smaller than the multipliers using BP identification within both the full-sample in Panel A as well as within the subsample in Panel B. The multiplier estimates for Finland derived from BP identification are also significantly different from zero. The estimates from CK identification are, however, not significantly different from zero with the chosen confidence band.[34]

In Panel B of Table 8 the fiscal multipliers are large especially with BP estimation.

[32] Canadian dollar was fixed between 1962-1970. Our instrument data starts from 1969Q1.

[33] The sample starts a few years earlier as Finnish Markka was integrated to the European Monetary System already in 1996. Note that before this period the exchange rate policy in Finland changed few times. While the exchange rate was most of the time fixed (pegged within a band), it was devalued (and revalued) repeatedly.

[34] One additional reason for this is probably that in CK there is one more structural parameter which is estimated and this results in more variation in the bootstrapped intervals. This can be seen, for example, in Figures 7 and 8 where the confidence intervals are larger at the initial impact for CK than they are for BP.

However, they are quite large also with CK identification even though these estimates are not significant even at the chosen significance level. This could be since the exchange rate regime is different compared with the full-sample which is a combination of several regimes. After all, if one again considers just the Mundell-Fleming model, one would expect that a small open economy with a fixed exchange rate reacts to fiscal policy. For a small country like Finland, the exchange rate is effectively fixed in a large monetary union. More so, the central bank of Finland no longer conducts independent domestic monetary policy. Therefore, the effects of fiscal shocks in Finland are not necessarily offset by the reactions of the European Central Bank.

Compared with earlier studies, the results on government spending multipliers are quite similar. For example, Virkola (2014) finds quite similar cumulative fiscal multipliers for Finland in the Euro regime period as well as for the full sample with Blanchard and Perotti (2002) identification. The results Lehmus (2014) reports are not directly comparable with our results as he does not report cumulative multipliers. Anyway the impulse response of output to government spending shock is somewhat larger at the 4 quarter and 8 quarter horizons than the ones we produce. This might be because of the different sample or also, for example, since Lehmus (2014) uses dummies for the recession at the beginning of the 1990s.

For Canada we are not aware of recent Blanchard and Perotti (2002) or Caldara and Kamps (2017) style estimates of cumulative fiscal multipliers that would be straightforwardly comparable with our estimates. Perotti (2005) reports also cumulative multiplier for sub-periods of his whole sample but he does not report cumulative multiplier for the whole sample period. Compared to his findings our BP_g identification scheme fiscal multiplier estimate for the whole sample is quite near the average of his estimates for the pre 1980 and post 1980 samples. But our estimate for the 1987-2013 period is notable larger than his estimate for the post 1980 period. The scaled impulse responses for the whole sample are shown for the BP_g identification scheme in Perotti (2005). Compared to our results the impulse response of output to a spending shock is somewhat higher at 4 quarters and 8 quarters. The impulse response of spending to a spending shock is somewhat smaller at 4 quarters and at 8 quarters than our responses. For example, the different time period and that Perotti's (2005) baseline model has 5 endogenous variables, might explain these differences.

Studying a panel of countries with the VAR framework Ilzetzki, Mendoza and Végh (2013) find support for the predictions of the Mundell-Fleming model. Similarly, we interpret that overall these multipliers are in line with the predictions of the Mundell-Fleming model. Note that we consider Finland and Canada as open economies. Contrary, Ilzetzki, Mendoza and Végh (2013) bundle these countries into a panel of closed economies or consider them open only for a certain period. The consensus anyway is the these countries are small open economies. Moreover, we see that the exchange rate of Canada is closer to flexible than fixed during the 1970-2018 period.

Panel A: Full-sample												
	BP_g			BP_r			CK_g			CK_r		
	$h=0$	$h=4$	$h=8$	$h=0$	$h=4$	$h=8$	$h=0$	$h=4$	$h=8$	$h=0$	$h=4$	$h=8$
IRF to g	1	1.18	1.08	1	1.16	1.04	1	1.22	1.18	1	1.2	1.11
	[1;1]	[1.07;1.36]	[0.9;1.39]	[1;1]	[1.05;1.35]	[0.85;1.34]	[1;1]	[1.09;1.4]	[0.96;1.49]	[1;1]	[1.07;1.38]	[0.9;1.41]
IRF to x	0.4	0.16	0.26	0.14	-0.39	-0.25	0.98	1.09	1.11	0.54	0.15	0.24
	[0.19; 0.61]	[-0.25; 0.91]	[-0.46; 1.26]	[-0.02; 0.36]	[-0.73; 0.33]	[-0.94; 0.78]	[0.28; 1.5]	[0.09; 2.03]	[0; 2.17]	[0.01; 0.93]	[-0.5; 1.02]	[-0.62; 1.38]
Cumulative multiplier	0.4	0.12	0.13	0.14	-0.32	-0.34	0.98	0.9	0.9	0.54	0.18	0.15
	[0.19; 0.61]	[-0.18; 0.55]	[-0.27; 0.7]	[-0.02; 0.36]	[-0.55; 0.12]	[-0.69; 0.25]	[0.28; 1.5]	[0.03; 1.57]	[0.04; 1.59]	[0.01; 0.93]	[-0.43; 0.78]	[-0.48; 0.83]
Panel B: 1987Q1-2013Q4												
	BP_g			BP_r			CK_g			CK_r		
	$h=0$	$h=4$	$h=8$	$h=0$	$h=4$	$h=8$	$h=0$	$h=4$	$h=8$	$h=0$	$h=4$	$h=8$
IRF to g	1	0.94	0.91	1	1.01	0.96	1	0.94	0.83	1	0.98	0.86
	[1;1]	[0.75;1.25]	[0.68;1.25]	[1;1]	[0.79;1.32]	[0.69;1.3]	[1;1]	[0.7;1.31]	[0.54;1.28]	[1;1]	[0.75;1.34]	[0.57;1.28]
IRF to x	0.84	1.23	0.11	0.44	0.42	-0.76	0.21	0.38	-0.5	-0.11	-0.22	-1.13
	[0.53; 1.12]	[0.29; 2.22]	[-0.87; 1.18]	[-0.02; 0.7]	[-0.73; 1.32]	[-1.82; 0.18]	[-0.93; 1.65]	[-1.15; 2.45]	[-1.84; 1.35]	[-0.89; 0.66]	[-1.37; 1.11]	[-2.24; 0.14]
Cumulative multiplier	0.84	1	0.85	0.44	0.29	0.03	0.21	0.09	0.01	-0.11	-0.44	-0.6
	[0.53; 1.12]	[0.31; 1.62]	[0.03; 1.6]	[-0.02; 0.7]	[-0.6; 0.77]	[-0.97; 0.68]	[-0.93; 1.65]	[-1.4; 2.14]	[-1.49; 2]	[-0.89; 0.66]	[-1.48; 0.59]	[-1.78; 0.5]
Panel C: 1987Q1-2013Q4 w/ foresight g and x												
	BP_g			BP_r			CK_g			CK_r		
	$h=0$	$h=4$	$h=8$	$h=0$	$h=4$	$h=8$	$h=0$	$h=4$	$h=8$	$h=0$	$h=4$	$h=8$
IRF to g	1	0.69	0.81	1	0.78	0.9	1	0.71	0.77	1	0.78	0.82
	[1;1]	[0.55;1.09]	[0.61;1.29]	[1;1]	[0.63;1.22]	[0.66;1.4]	[1;1]	[0.49;1.23]	[0.47;1.38]	[1;1]	[0.59;1.2]	[0.52;1.36]
IRF to x	0.88	0.46	-0.43	0.79	0.12	-0.99	0.2	-0.22	-0.72	-0.03	-0.68	-1.29
	[0.44; 1.12]	[-0.72; 1.59]	[-1.41; 1.03]	[0.14; 1.02]	[-1.27; 1.17]	[-2.04; 0.45]	[-1.07; 1.36]	[-2.01; 1.44]	[-2.32; 0.91]	[-1.05; 0.76]	[-2.16; 0.6]	[-2.65; 0.13]
Cumulative multiplier	0.88	0.65	0.33	0.79	0.36	-0.07	0.2	-0.38	-0.44	-0.03	-0.83	-0.94
	[0.44; 1.12]	[-0.29; 1.42]	[-0.63; 1.32]	[0.14; 1.02]	[-0.83; 1.09]	[-1.16; 0.86]	[-1.07; 1.36]	[-1.97; 1.5]	[-1.95; 1.27]	[-1.05; 0.76]	[-2.11; 0.49]	[-2.27; 0.33]

Table 7: Government spending multipliers, Canada.

Notes: Each panel reports results for the 4 different identification schemes. BP_g *and* BP_r *refers to Blanchard and Perotti (2002) identification with government spending first (g) and with net revenue first (r) respectively.* CK_g *and* CK_r *refers to Caldara and Kamps (2017) style identification with government spending first (g) and with net revenue first (r) respectively. For all identification schemes the panels report the values of the impulse responses of government spending and output to a shock in government spending at a certain period after the shock and a cumulative multiplier derived from these impulses at the corresponding horizon (0, 4 and 8 quarters). The 68% confidence intervals (Jentsch and Lunsford (2019) residual-based moving block bootstrap with 2000 draws) are given in brackets.*

Panel A: Full-sample

	BP_g			BP_r			CK_g			CK_r		
	$h=0$	$h=4$	$h=8$	$h=0$	$h=4$	$h=8$	$h=0$	$h=4$	$h=8$	$h=0$	$h=4$	$h=8$
IRF to g	1	0.82	0.65	1	0.81	0.66	1	0.74	0.52	1	0.73	0.53
	[1;1]	[0.67;1.11]	[0.42;0.94]	[1;1]	[0.63;1.06]	[0.4;0.92]	[1;1]	[0.55;1.01]	[0.23;0.78]	[1;1]	[0.53;0.98]	[0.25;0.78]
IRF to x	0.28	0.73	0.85	0.37	0.79	0.94	-0.45	-0.42	-0.34	-0.33	-0.34	-0.21
	[0.15; 0.47]	[0.13; 1.64]	[-0.23; 2.23]	[0.2; 0.56]	[0.17; 1.61]	[-0.2; 2.19]	[-1.16; -0.09]	[-1.49; 0.46]	[-1.85; 1.03]	[-0.8; -0.01]	[-1.13; 0.45]	[-1.53; 1.11]
Cumulative multiplier	0.28	0.57	0.81	0.37	0.65	0.9	-0.45	-0.55	-0.59	-0.33	-0.45	-0.46
	[0.15; 0.47]	[0.18; 1.12]	[0.1; 1.72]	[0.2; 0.56]	[0.23; 1.13]	[0.13; 1.73]	[-1.16; -0.09]	[-1.57; 0.09]	[-2.29; 0.48]	[-0.8; -0.01]	[-1.17; 0.13]	[-1.75; 0.52]

Panel B: 1995Q2-2018Q4

	BP_g			BP_r			CK_g			CK_r		
	$h=0$	$h=4$	$h=8$	$h=0$	$h=4$	$h=8$	$h=0$	$h=4$	$h=8$	$h=0$	$h=4$	$h=8$
IRF to g	1	0.19	0.05	1	0.19	0.05	1	0.12	0	1	0.13	0
	[1;1]	[0.07;0.32]	[-0.07;0.13]	[1;1]	[0.07;0.33]	[-0.07;0.13]	[1;1]	[-0.09;0.26]	[-0.17;0.07]	[1;1]	[0;0.27]	[-0.14;0.08]
IRF to x	0.47	1.42	1.88	0.48	1.42	1.88	-0.11	0.55	1.25	-0.07	0.59	1.29
	[0.26; 0.66]	[0.58; 2.17]	[0.5; 2.56]	[0.2; 0.65]	[0.52; 2]	[0.47; 2.44]	[-1.05; 0.17]	[-1.07; 1.24]	[-0.52; 1.86]	[-0.58; 0.24]	[-0.54; 1.28]	[-0.22; 1.98]
Cumulative multiplier	0.47	2.02	4.96	0.48	2.03	4.96	-0.11	0.38	2.5	-0.07	0.44	2.58
	[0.26; 0.66]	[0.95; 2.96]	[2.34; 6.85]	[0.2; 0.65]	[0.73; 2.73]	[2.02; 6.38]	[-1.05; 0.17]	[-2.74; 1.47]	[-3.07; 5.11]	[-0.58; 0.24]	[-1.48; 1.38]	[-2.4; 4.45]

Table 8: Government spending multipliers, Finland.

Notes: Each panel reports results for the 4 different identification schemes. BP_g *and* BP_r *refers to Blanchard and Perotti (2002) identification with government spending first (g) and with net revenue first (r) respectively.* CK_g *and* CK_r *refers to Caldara and Kamps (2017) style identification with government spending first (g) and with net revenue first (r) respectively. For all identification schemes the panels report the values of the impulse responses of government spending and output to a shock in government spending at a certain period after the shock and a cumulative multiplier derived from these impulses at the corresponding horizon (0, 4 and 8 quarters). The 68% confidence intervals (Jentsch and Lunsford (2019) residual-based moving block bootstrap with 2000 draws) are given in brackets.*

Appendix H - Data Appendix

In Figure 17 we show the forecast errors for the largest trading partners of Finland and the associated weights (share of total exports) used to construct the instrument for Finland. Column 1-6 represent each one of the largest trading partners. Column 7 represents all the other countries (pooled). The first row depicts the GDP forecast error and the second row depicts the import forecast errors. The forecast errors are log-differences. The third row depicts the export share of each country which is used as the weight when constructing the instrument. The sharp increase in the export share variable of other countries in row 3 column 7 is due to an increase in country coverage in the trading partner data.

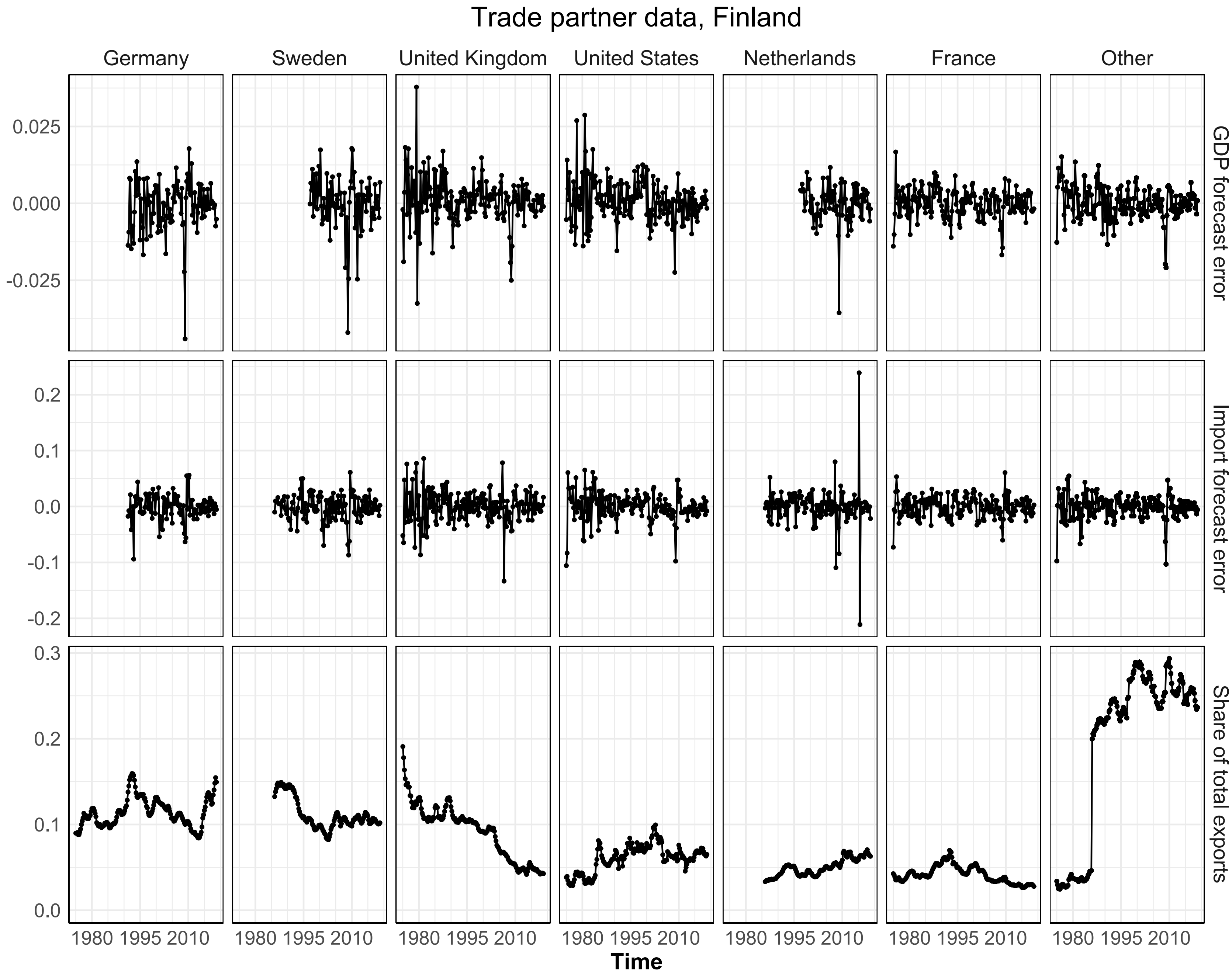

Figure 17: Trading partner data used in the construction of our instrument, Finland

Notes: In this figure we show the forecast errors for the largest trading partners of Finland and the associated weights (share of total exports) used to construct the instrument for Finland. Column 1-6 represent each one of the largest trading partners. Column 7 represents all the other countries (pooled). The first row depicts the GDP forecast error and the second row depicts the import forecast errors. The forecast errors are log-differences. The third row depicts the export share of each country which is used as the weight when constructing the instrument. The sharp increase in the export share variable of other countries in row 3 column 7 is due to an increase in country coverage in the trading partner data.