Comparing Two Methods for the Identification of News Shocks

Paul Beaudry, Franck Portier,
and Atılım Seymen

ZEW

Zentrum für Europäische
Wirtschaftsforschung GmbH

Centre for European
Economic Research

Discussion Paper No. 13-110

Comparing Two Methods for the Identification of News Shocks

Paul Beaudry, Franck Portier,
and Atılım Seymen

Comparing Two Methods
for the Identification of News Shocks[*]

Paul Beaudry[†]
Franck Portier[‡]
Atılım Seymen[§]

December 13, 2013

Abstract

Recent empirical literature delivered, based on different structural VAR approaches, controversial results concerning the role of anticipated technology—news—shocks in business cycle fluctuations. We deal with this controversy and investigate (i) the extent to which two prominent structural VAR approaches can be useful in recuperating news shock dynamics from artificially generated data in general and (ii) why and to what extent these SVAR approaches differ in the results they deliver in particular. Thereby, we provide several insights for the users of both VAR techniques with small samples in practice.

JEL classification: C32, E32

Keywords: News Shocks, Structural VAR, Identification

1 Introduction

Starting with the study of Beaudry and Portier (2004), there has been renewed interest in the idea that business cycles might mainly arise due to agents' difficulties in anticipating future developments. The term "news shock" which has been coined in this context refers to information on future fundamentals with no impact on current fundamentals. While this definition of news shock is rather general, the technology interpretation of news shocks gained particular popularity due to the empirical findings of Beaudry and Portier (2006, henceforth BP) and Beaudry and Lucke (2010), among others. This line of the literature found—based on structural vector autoregression (SVAR) techniques—that anticipated technology shocks have been a significant driving force of the business cycles in the US.

A serious challenge to the foregoing finding came recently from Barsky and Sims (2011, henceforth BS) who work with a different SVAR identification scheme than the former authors. The findings of BS attribute much less weight to technological news shocks in the forecast error variance of activity variables such as consumption, output and hours worked over a 2–3 year horizon than BP's findings.[1] The subsequent controversy is intriguing, because both sides of the discussion work with SVAR techniques and use similar data. Moreover, it has important implications as to how news shocks must be modeled in a theoretical framework. While the empirical findings of Beaudry and Portier (2006) and followers are hard to reconcile with the standard neoclassical setting, as has been discussed by Beaudry and Portier (2004, 2007), the empirical results of Barsky and Sims are in line with the implications of the standard neoclassical framework.

The aim of this study is to investigate the root of the difference in the findings of BP and BS. Furthermore, we want to examine whether SVAR techniques are in general appropriate for examining the impact of news shocks, since the ability of SVARs to recuperate structural shocks has often been questioned in the literature. The bottom line of the discussion is that the performance of SVAR identification schemes is highly dependent on the properties of the underlying data generating process.[2] A number of issues are known to play an important role in the success/failure of SVAR identification. Notably, truncation bias—that a finite-ordered VAR is only an approximation of the "true" dynamics—, small sample bias, relative volatility of shocks and the correct transformation of variables can have an impact on the quality of empirical findings that rely on SVARs. A further issue that becomes relevant particularly

---

[1]It must be noted, on the other hand, that both BP and BS obtain a significant weight of technological news shocks in the TFP forecast error variance in the long run.

[2]See Cooley and Dwyer (1998), Erceg, Guerrieri, and Gust (2005), Christiano, Eichenbaum, and Vigfusson (2007) and Chari, Kehoe, and McGrattan (2008), among others.

when one wants to estimate news shocks is the so-called non-invertibility problem, i.e., that there can be no one-to-one mapping between the structural shocks of the true economy and the ones estimated using SVAR techniques.[3]

In order to address the above listed issues, we carry out Monte Carlo simulations using artificial data that are generated with two different calibrated theoretical models serving as data generating process (DGP). Note that this is what Beaudry and Lucke (2010) (in a web-appendix) or Barsky and Sims (2011) do in order to verify that the identification strategies they use are capable of recuperating "true" impulse responses. While in both cases the authors can show an example of their empirical strategy being plausible for dealing with the question of interest, as noted above, the question of whether technological news shocks are an important driver of short-term cyclical fluctuations is still answered both positively and negatively. Moreover, both Beaudry and Lucke (2010) and Barsky and Sims (2011) investigate only whether their own identification schemes function well given a particular DGP, without discussing the performance of alternative identification schemes.

To be fair to both BP and BS approaches, we work with two alternative DGPs as our laboratory. We choose the first DGP to be a slightly different version of the standard real business cycle (RBC) model with real frictions—habit formation in consumption and adjustment cost function in investment—that is used by Barsky and Sims (2011) in order to evaluate the goodness of their own identification scheme, whereas the second DGP we use is the one-sector model of Jaimovich and Rebelo (2009, henceforth JR). While the latter DGP contains the adjustment cost function for investment as well, it differs from the former DGP in comprising two further elements that are crucial for it to generate the well-known comovements among output, consumption, investment and hours worked after the occurence of a news shock. Those additional elements are variable capital utilization and the so-called GHH preferences of Greenwood, Hercowitz, and Huffman (1988). By means of our simulations, we aim to show the extent to which both identification schemes differ in recuperating news shocks and the corresponding IRFs under both DGPs. This should help understand the root of the above described empirical controversy.

Our main result is that, when the same reduced-form underlies the estimation, BP and BS models deliver quite similar results regardless of the properties of the underlying true DGP. What matters most for the success of both identification schemes is what type of

---

[3]See, e.g., Fève, Matheron, and Sahuc (2009), Leeper, Walker, and Yang (2011), **?** and Sims (2012). As Sims (2012) shows, formal existence of non-invertibility does not necessarily imply that SVAR models are useless. Including enough forward-looking variables in the model might lead to a good approximation of the true dynamics by SVARs.

information is included in the variables comprised by the VAR. This issue has been stressed by Beaudry and Portier (2013) recently, who show that the four-variable VAR of Barsky and Sims (2011) obtains results similar to the ones reported by Beaudry and Portier (2006) when estimated with the same variables. In this paper, we show that with the same trivariate or four-variable structure, the BP and BS schemes arrive at similar conclusions, but the results may differ substantially across trivariate and four-variable models in general.

We compare trivariate and four-variable versions of BP and BS models and find that the models we consider perform fairly well in recuperating the true IRFs as well as the true shocks from the data in most cases. This is true in spite of the non-fundamentalness problem that arises when news shocks drive the system. From this perspective, our conclusions are similar to the conclusions of Sims (2012). We further provide insights on the use of SVARs with small samples. Given that sample sizes have certain limits in practice, our results should be helpful for practitioners in their choice of the lag order and cointegration specification of the VARs. Our findings also address the role of the relative size of news shocks, the anticipation lead and measurement problems related to TFP in the context of news shock identification.

The paper is organized as follows. The next section presents the BP and BS approaches to news shock identification. The structure of the simulation exercises is described in Section 3. The results of simulations are discussed in Section 4. Section 5 concludes.

2 Identification of News Shocks: Two Alternative Approaches

2.1 The Reduced Form

In order to facilitate a healthy comparison, identification with both BP and BS approaches should build on the same reduced form estimation. One difference, of no major importance as we discuss later, between the original BP and BS studies is that Barsky and Sims (2011) prefer a level estimation, whereas BP estimate vector error correction models (VECMs). Although both approaches have their merits and asymptotically lead to the same result, we initially estimate a VECM with K variables and a cointegration rank of $K - 1$ for both BP and BS frameworks for the sake of comparibility. Note that this cointegration rank is consistent with the structure of the DGPs we work with as will be seen later. For any empirical model we consider, the variables are collected in vector x_t, and the first vari-

able is always log total factor productivity (a_t). We include either log stock price (sp_t) or log consumption (c_t) as the second, forward-looking, variable in the forthcoming simulation exercises. This system is extended with log hours worked (n_t) in the trivariate case and further extended with log output (y_t) in the four-variable case. The (reduced form) Wald representation of any system we estimate is given by

$$\Delta x_t = C\left(L\right)u_t, \tag{1}$$

where Δ is the first difference operator, x_t is a K-dimensional vector of endogenous variables, $C\left(L\right) = I + \Sigma_{i=1}^{\infty}C_iL^i$, the reduced form innovation vector u_t has a non-singular covariance matrix, and deterministic terms are skipped for ease of presentation.

2.2   Identification in Beaudry and Portier (2006)

In our context, the aim of an econometrician is to affiliate the reduced form with the structural form, i.e., with

$$\Delta x_t = \Gamma\left(L\right)\varepsilon_t, \tag{2}$$

where $\Gamma\left(L\right) = \Sigma_{i=0}^{\infty}\Gamma_iL^i$ and the covariance matrix of the structural shocks, ε_t, is set to be identity matrix. Following BP, we employ the following restrictions in order to identify surprise and anticipated technology shocks.

Bivariate Framework   The starting point of BP is a bivariate model where the crucial restriction for identifying news shocks is the zero-restriction in the short-run multiplier matrix,

$$\Gamma_0 = \begin{bmatrix} * & 0 \\ * & * \end{bmatrix}, \tag{3}$$

i.e., the $(1,2)$ element of Γ_0 is set to zero. This implies that the second shock of the bivariate system has no contemporaneous impact on the TFP, but is not restricted in terms of its impact after the initial period. The second shock hence stands for a news shock.

Trivariate Framework   The trivariate framework needs three restrictions for an exact identification together with the ones on the covariance matrix of the structural shocks. The

4

additional three restrictions in the BP framework are given by

$$\Gamma_0 = \begin{bmatrix} * & 0 & 0 \\ * & * & * \\ * & * & * \end{bmatrix} \text{ and } \Gamma(1) = \begin{bmatrix} * & * & 0 \\ * & * & * \\ * & * & * \end{bmatrix}. \tag{4}$$

Thus, only the first shock – the surprise TFP shock – can have an impact effect on TFP. The news shock, on the other hand, does not alter the TFP at impact, but can have a long-run effect on it. Note that both surprise and anticipated technology shocks can have a long-run impact on TFP according to this specification. Finally, the third shock can represent a measurement error or a linear combination of other structural shocks in the data generating process.

Four-Variable Framework The four-variable framework is quite similar to the trivariate one with the following crucial restrictions:

$$\Gamma_0 = \begin{bmatrix} * & 0 & 0 & 0 \\ * & * & * & 0 \\ * & * & * & * \\ * & * & * & * \end{bmatrix} \text{ and } \Gamma(1) = \begin{bmatrix} * & * & 0 & 0 \\ * & * & * & * \\ * & * & * & * \\ * & * & * & * \end{bmatrix}. \tag{5}$$

The first and second shocks are again labelled surprise TFP and news, respectively, while we do not comment on the third and fourth shocks. There are again two shocks with a permanent impact on TFP and the last two shocks could represent a measurement error or a linear combination of structural shocks other than technology shocks in the data generating process. Although we arbitrarily set the $(2,4)$ element of Γ_0 to zero, notice that whether the $(2,3),(2,4)$ or $(3,4)$ element of Γ_0 is set to zero does not make a difference in terms of the identification of surprise TFP or news shocks within this structure.

2.3 Identification in Barsky and Sims (2011)

The identification scheme proposed by Barsky and Sims is based on the same assumption as to surprise technology shocks: these are the only driver of TFP at the impact period. BS conjecture furthermore, similar to BP in the trivariate and four-variable models, that there are only two sources of shocks that can impact on TFP, namely surprise TFP and news shocks. Further shocks in a system with at least three variables are not supposed to effect

TFP at any time. Thus, the condition

$$\Omega_{1,1}(h) + \Omega_{1,2}(h) = 1 \ \forall h \tag{6}$$

must be valid, where $\Omega_{i,j}(h)$ stands for the share of the forecast error variance (FEV) of variable i attributable to structural shock j at horizon h, and $i = 1$ referring to TFP in our framework. Having identified the first shock as a surprise technology shock, BS choose the second shock in the VAR such that it explains the remaining TFP FEV over the horizon 1 to H as much as possible. The authors solve the following maximization problem to this end:

$$\gamma^* = \arg\max \sum_{h=0}^{H} \Omega_{1,2}(h)$$

$$\text{s.t.}$$

$$\tilde{A}_0(1,j) = 0 \ \forall j > 1$$

$$\gamma(1,1) = 0$$

$$\gamma'\gamma = 1$$

Here $\tilde{A}_0$ is the lower triangular Cholesky factor of the covariance matrix of reduced form innovations u_t. γ is an orthonormal vector such that $\tilde{A}_0\gamma$ stands for the impact response of the variables in the system to the second shock in the VAR, the news shock.

2.4 Comparison of BP and BS schemes

BP and BS schemes identify the surprise technology shock identically but may differ in the identification of the news shock. With respect to the latter shock, both approaches share the restriction that it has no impact effect on TFP. The BS approach furthermore embraces the solution of a maximization problem, whereas the BP approach imposes that no other shock than surprise and anticipated technology can have a long-run impact on TFP. Note that the BS approach tries to fulfil the restriction in (6) over $h = 1, \ldots, H$ as precisely as possible, while the BP approach imposes the same condition only at $h = \infty$ but precisely. Both approaches give the same result in a bivariate system by construction. As Barsky and Sims (2011) emphasize, the BS and BP approaches would asymptotically lead to the same result if the conditions for the BP identification to be valid were satisfied. Furthermore,

the performance of both schemes in terms of news shock identification is dependent on how successfully they can recuperate the surprise technology shock.

Finally, the choice of the horizon over which the BS maximization problem is solved matters for the performance of the BS scheme. This issue has recently been pointed out by Beaudry, Nam, and Wang (2011). A criteria by means of which the value of the parameter H can be optimized when working with real data does, however, not exist. The sensitivity of BS results to the value of H is therefore one of the issues we investigate below.

3   Structure of Simulation Exercises

In this section we outline the design of the simulation exercises. What we do below in essence is to generate artificial data with calibrated theoretical models and use these to compare the performance of the identification schemes we described in the foregoing section. In the following, we first discuss the two DGPs that underlie our simulations. This is followed by the description of benchmark simulation as well as a number of alternative simulations. The last subsection lists the tools that are used for evaluating the SVAR identification schemes in this study.

3.1   DGPs

As mentioned in the introduction, our analysis builds on two alternative models serving as DGP—an RBC model real frictions and the one-sector model of Jaimovich and Rebelo (2009). We feed both models with the same two exogenous processes through which they are exposed to stochastic shocks.[4] Notably, we assume in both models that log TFP follows the process

$$\ln A_t = g_A + \ln A_{t-1} + \varepsilon_{1,t} + \varepsilon_{2,t-j}, \tag{7}$$

$\varepsilon_{1,t}$ and $\varepsilon_{2,t-j}$ respectively being surprise and anticipated (news) technology shocks, j indicating the anticipation lead, and g_A indicating the trend growth rate of TFP. Both models contain furthermore shocks to the share of goverment spending in GDP of which law of motion is modelled as

$$\ln g_t = (1-\rho)\ln \bar{g} + \rho \ln g_{t-1} + \varepsilon_{3,t} + \varepsilon_{4,t-j}, \tag{8}$$

---

[4]Note that we substitute the investment-specific technology (IST) shock in the original model of Jaimovich and Rebelo (2009) with the government spending shock here in order to ensure comparibility of the results across our simulations.

$\varepsilon_{3,t}$ and $\varepsilon_{4,t-j}$ standing for surprise and anticipated shocks to goverment spending share, $\bar{g}$ being the steady state government spending share, and $0 \leq \rho < 1$ determining the persistence of shocks on government spending. TFP is thus the only source of a stochastic trend in both models, while government spending follows a stationary process. As long as not stated otherwise, the standard deviations of $\varepsilon_{1,t}, \varepsilon_{2,t}, \varepsilon_{3,t}$ and $\varepsilon_{4,t}$ are respectively set to 0.66, 0.66, 0.25 and 0.25 in the simulations.

3.1.1 DGP 1: RBC Model with Real Frictions

The RBC model with real frictions used by Sims (2009), which is very similar to the one considered by Barsky and Sims (2011), as a DGP to evaluate the performance of their identification scheme is the first DGP we work with in this study. The model contains habit formation in consumption, labor/leisure choice, investment adjustment costs in the capital accumulation equation, a Cobb-Douglas production function with constant returns to scale, and government spending of which GDP share is subject to stochastic shocks mentioned above. The following planner's problem describes together with (7) and (8) the model:[5]

$$\max E_0 \sum_{t=0}^{\infty} \beta^t \left(\ln \left(C_t - bC_{t-1} \right) - \psi \frac{N_t^{1+1/\eta}}{1+1/\eta} \right)$$

$$\text{s.t.}$$

$$K_t = (1-\delta) K_{t-1} + \left[1 - \phi \left(\frac{I_t}{I_{t-1}} \right) \right] I_t$$

$$Y_t = A_t K_{t-1}^{\alpha} N_t^{1-\alpha}$$

$$Y_t = C_t + I_t + G_t$$

$$G_t = g_t Y_t$$

Following Barsky and Sims (2011), the model is calibrated with the following parameter values: $\beta = 0.99$, $b = 0.8$, $\psi = 1$, $\eta = 1$, $\delta = 0.025$, $\alpha = 0.33$, $\bar{g} = 0.2$, $g_A = 0.0025$, $\rho = 0.95$. Moreover, $\phi(\cdot) = \phi'(\cdot) = 0$ and $\phi''(\cdot) = 0.3$.

---

[5]Note that the time convention corresponding to K_t is different here than in the foregoing papers for notational convenience.

3.1.2 DGP 2: One-Sector Model of Jaimovich and Rebelo (2009)

Jaimovich and Rebelo develop a framework where news shocks are able to generate co-movement that is typically observed across macroeconomic quantities over the business cycle. Their one-sector model differs in two aspects from the foregoing one: it comprises variable capital utilization and GHH preferences where consumption and hours worked decisions are non-time-seperable. Moreover, JR do not include a government spending variable of the type above but let investment in the economy-wide budget constraint to be subject to IST shocks. We discard IST shocks from the model and assume the same steady state growth rate for TFP as in DGP 1 in our forthcoming simulations for the sake of subjecting both DGPs to the same type of exogenous shocks. The following planner's problem describes together with (7) and (8) the second DGP we use in our simulations:

$$\max E_0 \sum_{t=0}^{\infty} \beta^t \frac{\left(C_t - \psi N_t^\theta X_t\right)^{1-\sigma} - 1}{1 - \sigma}$$

$$\text{s.t.}$$

$$X_t = C_t^\gamma X_{t-1}^{1-\gamma}$$

$$K_t = (1 - \delta(u_t)) K_{t-1} + \left[1 - \phi\left(\frac{I_t}{I_{t-1}}\right)\right] I_t$$

$$Y_t = A_t \left(u_t K_{t-1}\right)^\alpha N_t^{1-\alpha}$$

$$Y_t = C_t + I_t + G_t$$

$$G_t = g_t Y_t$$

In the benchmark simulations, we calibrate the parameters of the model with the values provided by JR: $\theta = 1.4$, $\beta = 0.985$, $\alpha = 0.36$, $\phi''(g_A) = 1.3$ and $\delta''(u) u/\delta'(u) = 0.15$. Furthermore, we set $\gamma = 0.005$. Note that the parameter values chosen for DGP 2 thus differ to some extent from the corresponding ones chosen for DGP 1. In particular, the implied steady state value for $\delta(u)$ in the JR calibration is more than four times as large as the (more standard) one chosen by BS. The second derivative of the investment adjustment cost function at the steady state also assumes a much higher value in the JR model than in the previous DGP. Slight differences exist also as to the values of the discount parameter of households and the share of capital in the production function. We discuss the impact of these differences on our results in the next section.

3.1.3 Variables Used in the Simulations

We report results for only trivariate and four-variable VAR models below. The bivariate system is uninteresting for two reasons. First, BP and BS schemes produce the same result in the bivariate case by construction. Thus, there is no need for comparison across the two identification approaches for that case. Second, there is an obvious misspecification in a bivariate model given the four types of shocks that enter our DGPs. The government spending shocks cannot find a place in a bivariate model as long as the model is supposed to identify surprise and anticipated technology shocks correctly.[6]

As mentioned in the previous section, the first variable in the VARs is always TFP in the following. Furthermore, the third and fourth variables in the VARs we consider are set to be log hours worked and log output. As to the second variable, we consider two alternatives. In the first case, we take it to be a measure of stock price, defined as

$$sp_t = q_t k_{t-1},$$

where q_t stands for the relative price of capital defined by

$$q_t = \frac{\eta_t}{\lambda_t},$$

η_t and λ_t are the Lagrange multipliers corresponding to the capital accumulation equation and economy-wide resource constraint in both DGPs, respectively. Alternatively, we try out consumption as the second variable in the VARs.

3.2 Simulations

In Section 4, we discuss the findings from our simulations. In each simulation exercise, 500 different data sets are generated using one of the foregoing DGPs where each data set contains 200 observations. More specifically, we generate 500 observations starting from the steady state of the underlying DGP and discard the first 300 observations afterwards for each data set. The calibration that we use implies that we thus produce data for 200 *quarters*. The benchmark anticipation lead for news shocks is set to two periods. The VAR lag order is set to four.

---

[6]Notice that misspecification might also be valid for trivariate models. However, when the aim is to identify surprise and anticipated technology as well as further demand shocks, the trivariate specification can still work fine if the VAR identifies the surprise technology and news shocks correctly and the surprise and anticipated government spending shocks as a linear combination.

We check the robustness of our conclusions from the benchmark simulations in a number of alternative exercises. The literature on the performance of SVARs mentioned in the introduction points to several factors that could have an impact on the quality of SVAR identification results. Notably, the following general characteristics are known to potentially affect SVAR performance:

- **Sample size:** Any statistical result is reliable only when it is based on a long enough sample, although there is no general rule as to the ideal minimum number of observations. Small samples are ominous when working with VARs due to geometrically increasing number of parameters to be estimated with increasing number of variables. We check the impact of sample size on SVAR results by decreasing the number of observations in the artificial data sets from 200 to 100 as well as increasing it to 1000.

- **Model specification**

  - **Lag order:** Finite order VARs that in practice typically underlie SVAR identification schemes provide only an approximation of "true" data generating processes which imply a VARMA representation for the variables. The so-called *truncation bias* may arise because of that approximation.[7] Note that, given the relatively short samples at hand most of the time, a high lag order can increase the estimation uncertainty. A lower lag order can, on the other hand, lead to inferior estimates due to omitted significant lags. We set the lag order to four in the benchmark simulation. In alternative exercises, we try out both the higher and lower lag orders of 6 and 2, respectively.

  - **Cointegration:** The handling of cointegration is another important aspect of VAR model specification. While both level VAR and VECM estimations are asymptotically fine, their performances may differ in small samples. Recall that Barsky and Sims build on a level VAR estimation, while Beaudry and Portier estimate VECMs in the original papers. We investigate the impact of an alternative way of handling cointegration by taking the cointegrating vectors to be known, i.e., we choose them to be exactly as in the DGP, compute the error-correction terms accordingly and estimate the remaining coefficients of the reduced-form VECM with OLS. This allows us to abstract from well-known problems regarding the estimation of cointegrated models. In both of the DGPs we consider,

---

[7]See, e.g., Cooley and Dwyer (1998), Erceg, Guerrieri, and Gust (2005) and Chari, Kehoe, and McGrattan (2008).

five variables—$\log Y_t, \log C_t, \log I_t, \log G_t$ and $\log K_t$—share a common stochastic trend given by $\log A_t^{1/(1-\alpha)}$. Note furthermore that hours worked is stationary. Thus, there must be two cointegrating relationships in trivariate models with $x_t = [\log A_t, \log C_t, \log N_t]$, given by

$$\beta = \begin{bmatrix} 1 & \alpha - 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{9}$$

For four-variable models with $x_t = [\log A_t, \log C_t, \log N_t, \log Y_t]$, three cointegrating vectors are given by

$$\beta = \begin{bmatrix} 1 & \alpha - 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & \alpha - 1 \end{bmatrix}. \tag{10}$$

For both trivariate and four-variable models, the second cointegrating vector has a unit entry corresponding to the stationary variable $\log N_t$ and zeros elsewhere, whereas the other cointegrating vectors are due to the fact that the generic growing variable reads $X_t = A_t^{1/(1-\alpha)} x_t$.

- **Relative size of shocks:** The literature mentioned in Footnote 2 on the identification performance of SVARs emphasizes the importance of the relative size of shocks as a factor contributing to their success/failure. In general, an increase in standard deviation of other shocks, such as government spending, creates higher noise in the data and can thus be expected to render the identification news shocks more difficult. Concerning our context, **?** have recently shown within an analytical framework how important the relative size of surprise and anticipated technology shocks are for the identification of the latter. Recall that we set $\sigma_{\varepsilon_1} = 0.66, \sigma_{\varepsilon_2} = 0.66, \sigma_{\varepsilon_3} = 0.25$ and $\sigma_{\varepsilon_4} = 0.25$ in the benchmark DGPs, where σ_{ε_i} stands for the standard deviation of $\varepsilon_{i,t}$ in terms of (7) and (8). We try out three alternatives; (1) $\sigma_{\varepsilon_1} = 0.33$, (2) $\sigma_{\varepsilon_2} = 0.33$ and (3) $\sigma_{\varepsilon_3} = \sigma_{\varepsilon_4} = 0.66$, where for each alternative we keep the standard deviations of the remaining shocks as in the benchmark case.

In addition to the hitherto listed factors that can play a role in the empirical success/failure of SVARs in general, further factors can be added to the list when it comes to the identification of *technological news* shocks. In our alternative simulations, we notably take the impact of the following factors into account:

- **Anticipation lead:** The average anticipation lead for news shocks is not easy to measure. While we choose a lead of two periods in our benchmark simulations, we try out four leads as an alternative. We do not work with longer anticipation leads, but it can intuitively be expected that identification of news shocks gets more difficult under longer anticipation leads. On the other hand, Schmitt-Grohé and Uribe (2009) report that changes in the stationary and nonstationary component of productivity are for the most part anticipated three quarters and one quarter in advance, respectively. Beaudry and Portier (2004) provide an estimate of one year, i.e. four quarters. Therefore, we believe, exercises with an anticipated lead of up to four quarters are enough to give an idea about the goodness of the identification schemes in this respect.

- **Measurement of TFP:** TFP is a crucial variable for the identification of news shocks. Measurement of TFP poses yet a challenge for statisticians in most cases. In order to account for this potential vulnerability, we consider two alternative exercises in this study.

  - **Noise when measuring TFP:** Both Beaudry and Portier (2004) and Jaimovich and Rebelo (2009) estimate theoretical models where the agents receive only noisy signals about the TFP dynamics in the future. Blanchard, L'Huillier, and Lorenzoni (2009) investigate the role of news shocks in macroeconomic fluctuations when agents receive information about the permanent component of TFP with some noise. The latter authors argue that news and noise shocks cannot be identified by means of SVARs when econometricians do not possess more information than agents. In order to account for noise in the data in a somewhat different context, namely that the noise exists only for the econometrician, we add an AR(1) noise term to the TFP growth rate in alternative simulations. The persistency parameter of the noise term is set to 0.5 and the standard deviation of the shock hitting that process is set to 0.33, i.e., one half the standard deviation of technological news shocks in our benchmark data generating processes. The signals that the econometricians receive about TFP developments are thus fairly noisy. Moreover, the persistency parameter of the noise term implies that the noise about TFP growth disappears largely after one year when working with quarterly data.

  - **Substituting TFP with labor productivity:** The availability of TFP is limited for many countries, particularly at quarterly frequency, so applied researchers

might use labor productivity as a proxy. Not least due to low variation in capital stock series, this may indeed be a good approximation. In the last exercise we play with this idea and substitute TFP with labor productivity in the empirical models. Note that this means not estimating the four-variable models now, since we would otherwise encounter a multicollinearity problem. Therefore, we do not include output, y_t, in the models directly in such an exercise. Yet, the trivariate models include it indirectly as a linear combination of labor productivity and hours.

Finally, we consider the impact of two further factors in our exercises:

- **The optimization horizon in the BS scheme:** Beaudry, Nam, and Wang (2011) show in a recent paper that the results from the BS scheme are sensitive to the length of the optimization horizon. Therefore, we investigate the impact of this parameter on the BS identification.

- **Changing the paratemer values in the DGP:** Some parameters in the DGP can be particularly crucial when identifying shocks by means of SVARs as Erceg, Guerrieri, and Gust (2005), e.g., point out. We have mentioned some differences between the calibration of DGP 1 and DGP 2 above. The sensitivity of our results to the calibration of DGP is therefore another aspect that we discuss in the next section. We do this by playing with several parameter values in the second DGP and bringing it closer to DGP 1 gradually.

3.3   Evaluation of the Identification Schemes

In our simulations, we focus on the performance of identification schemes with respect to impulse response functions as it became convention in the related literature. We present the true impulse response from the DGP together with the ones from the simulations corresponding to each empirical model graphically. To be more precise, we plot the median estimated response corresponding to each empirical model. We can thus establish how far the distance between the true response and median empirical responses is. Note that such a distance measure gives an idea about the success of identification schemes on average only, assuming that a large number of data sets are available. In practice, however, there is only one data set econometricians can work with. Hence, it is useful to also look at the dispersion of the empirical impulse responses around the true impulse response as we do in alternative

graphs below. Finally, we report the empirical distribution of correlation between the true and estimated news shock series for each simulation. Although this correlation can never be perfect in our framework due to the non-fundamentalness problem, a high correlation would very likely indicate that the problem is negligible in practice.

4 Results

4.1 Benchmark Simulations

4.1.1 DGP 1

Models with Stock Price as the Second Variable   In our first exercise with DGP 1, we use stock price as the second variable in the VARs. The upper panel of Figure 1 shows the true response of TFP, consumption, output and hours to a news shock together with the median impulse responses derived from the individual empirical models. According to these graphs, the performances of BP and BS identification schemes are very similar when they rely on the same (trivariate or four-variable) reduced-form VAR. In other words, only the information included in the variables that go into the VAR seems to be important for explaining differences in the performance of the identification schemes. Whereas BP and BS schemes perform similarly when they depend on the same reduced form, the performance of the trivariate and four-variable systems vary substantially. Although all models capture the IRFs qualitatively well, the median responses from the four-variable systems are by far much closer to the true response functions than the responses obtained with trivariate models. This indicates that having output in addition to the first three variables of the VAR brings a significant amount of new information for identification with the first DGP. Note furthermore that the median responses for all variables from the VARs are closer to the corresponding true response at shorter horizons, while the median response estimates deteriorate as the impulse response horizon increases.

Turning to the confidence intervals in the form of 0.05 and 0.95 percentile of the IRF distributions from the simulations, which are illustrated in the lower panel of Figure 1, we see that the intervals generated by the four-variable models successfully capture the true responses, whereas trivariate models fail in this respect. Note that the four-variable BP model performs slightly worse than its BS counterpart in terms of the lower bound confidence interval estimates. Moreover, the true responses are somewhat closer to the upper bound of the confidence intervals than to the lower bound, suggesting an overall downward bias in the

estimated impulse responses from the VARs.

Finally, recall that, if the identification schemes perform perfectly, they must obtain a correlation of one between true and empirical shocks. That case cannot arise due to the non-invertibility problem as mentioned above. Yet, the four-variable models are successful in generating a high correlation, i.e., one that is above 0.8, in most cases, as the distribution of the correlation between true and empirical news shocks, shown in Figure 2, suggests. On the other hand, the trivariate models fail seemingly in this respect. This confirms our foregoing conclusions: four-variable models are more successful in getting a high correlation and the BS model is slightly superior to the BP model in this context, but trivariate models seem not to include a sufficient amount of information for a satisfactory identification of shocks.

Models with Consumption as the Second Variable We now turn our attention to our alternative exercise with DGP 1, where stock price is substituted with consumption as the second variable in the VARs. According to the upper panel of Figure 3, all models perform rather similarly in getting the true impulse response on average. It is remarkable that the empirical models are rather successful in getting the form of the impulse response functions (IRFs) correct, even in the more challenging case of hours worked. Although the median IRFs point to a slight underestimation for all variables, the distances between true and median IRFs are not too large. As with the case where stock price is the second variable of the VAR, the IRF estimation performance of the VARs deteriorates at longer horizons such that the distance between the median responses from the VARs and the corresponding true response gets wider. Finally, the four-variable BP model performs slightly, indeed negligibly, worse than the other models considered, especially at shorter horizons of up to 10 quarters.

In the lower panel of of Figure 3, all models seem to generate confidence intervals that well cover the true response, although with a bit of downward bias. Again, the four-variable BP model performs slightly worse than the other models considered, particularly in terms of the lower bound confidence interval estimates.

As to the distribution of the correlation between true and empirical news shocks illustrated for each empirical model in Figure 4, the models are successful in generating a high correlation, i.e., one that is above 0.8, in most cases. The trivariate models are slightly more successful according to this measure than the four-variable models. In particular, there is almost a 50% likelihood that the four-variable BP model estimates an empirical news shock series that is correlated with the true series with a coefficient less than 0.8.

In comparison to the previous model where stock price was the second variable, the performance of the trivariate models improve substantially and four-variable models marginally

when consumption substitutes stock price in the VAR. This suggests on the one hand that consumption bears more valuable time series information than stock price for news-shock identification in our artificial world. On the other hand, when consumption is used in the four-variable model instead of stock price, this leads to a slight deterioration of the median impulse-response estimates. This must be due to the fact that consumption and output deliver similar time series information to a large extent and the marginal information that consumption brings in a four-variable system is therefore less than the information brought by stock price.

4.1.2 DGP 2

In this subsection, we report the results from our exercises which are carried out by using the second DGP. Before we start with the presentation of the results, however, a remark is in order. A comparison of the impulse responses shown in the following with the foregoing ones points to a significant difference between the two DGPs that we consider in this study: the second DGP generates much stronger fluctuations than the first DGP. Whereas the TFP responses across the two DGPs are the same by construction, we observe that the scales of the consumption, hours worked and output IRF graphs are much larger under the second DGP than the first one. This means that the sensitivity to shocks of the second economy is much higher than the first economy—not without consequences for the performance of the SVAR identification schemes as we discuss later.

Models with Stock Price as the Second Variable As with DGP 1, we start out our presentation of the results with DGP 2 where the VARs include stock price as the second variable. The corresponding IRF results are given in Figure 5. In the upper panel of the figure, both trivariate and four-variable models generate qualitatively good-looking median responses. Note that only the TFP impulse responses are comparable across DGP 1 and DGP 2. The comparison of the VARs with the stock price as the second variable shows that the performance of the BP and BS schemes is somewhat weaker under DGP 2, which in general leads to a higher variance of the endogenous variables as reflected by much stronger impulse response functions. In particular, the BP and BS schemes recuperate the impulse responses in the first 8 quarters rather well, while their performances start to worsen significantly from that horizon on.

Another conspicuous difference from the case with DGP 1 is that, under DGP 2, the BS identification scheme slightly overperforms the BP identification scheme. The reason behind

the relatively worse performance of the BP scheme is its dependence on the estimation of the long-run multiplier matrix $\Gamma(1)$ in equations (4) and (5). Note that the estimation of $\Gamma(1)$ becomes a more challenging task with data showing relatively higher variance under DGP 2. In contrast, good estimation of the first H moving-average coefficient matrices in the VAR is enough for successful identification with the BS scheme.[8] This insight explains why the performances of both the BP and BS schemes worsen under DGP 2, but the BP-scheme performance more than the BS-scheme performance.

Turning to confidence intervals in the lower panel of Figure 5, as under DGP 1, only four-variable schemes can generate confidence intervals that contain the true responses. It must be noted, however, that the true responses tend to be further closer to the upper bound than in the foregoing exercises. This suggests that the downward-bias in the estimation of IRFs is larger under DGP 2.

Finally, the histograms in Figure 6 confirm that four-variable models achieve better results than trivariate models with DGP 2 and stock price as the second variable in the VAR, BS schemes performing somewhat better than BP schemes in terms of both trivariate and four-variable models.

Models with Consumption as the Second Variable Figure 7 shows the IRFs under the second DGP, where consumption is the second variable in the VAR. As in the foregoing exercise, we observe that the four-variable models are much succesful than the trivariate models in coming close to the true response in median, and this *both* qualitatively and quantitatively. Similar to the previous results, both trivariate and four-variable BS models are more succesful in capturing the true responses than their BP counterparts quantitatively, although there is no qualitative difference across the IRFs corresponding to the BP and BS schemes.

In terms of the confidence intervals generated by the models, again, only four-variable models are able to generate intervals that include the true responses somewhat in the middle, whereas the trivariate models fail in this respect as can be seen from the lower panel of Figure 7.

Another striking difference to the hitherto simulation exercises relates to the distribution of the correlation between the true and empirical news shocks. As can be seen in Figure 8, the trivariate models can obtain a correlation coefficient above 0.8 in none of the 500 simulations that we carry out. In contrast, the four-variable models are much more succesful

---

in this respect, the BS scheme generating a high correlation much more often than the BP scheme.

4.1.3   Remarks

A few lessons from our benchmark simulations are worth pointing out before we turn our attention to alternative simulations and check the sensitivity of our hitherto findings. First, confirming previous results in the literature, we find that SVARs can be a useful tool for estimating structural shocks in general. Second, despite concerns raised regarding the non-invertibility problem in the context of the identification of news shocks in the literature, SVARs can lead to a satisfactory estimation of news shocks and their dynamic effects. Third, which information is included in the VAR is key for the performance of SVAR identification schemes. Fourth, the properties of the true DGP are decisive for the success of SVAR identification schemes. In case the true DGP implies a high variation of the variables in the system, it gets harder to identify structural shocks reliably with SVARs. Fifth, the trivariate and four-variable BP identification schemes are somewhat more sensitive to higher variation in the underlying data than the corresponding BS schemes.

With the foregoing benchmark simulation exercises, we could hardly detect any factor that can explain the discrepancy between the BP and BS results in the empirical literature. In the next subsection we discuss the impact of further issues we listed in subsection 3.2 on the success/failure of the identification schemes.[9]

4.2   Alternative Simulations

4.2.1   Changing the Parameter Values in the DGP

We have pointed out above that the two DGPs differ in two aspects: calibration of several common parameters and two model mechanisms—utility function specification and inclusion/exclusion of variable capacity utilization. Whereas the utility function of the household in DGP 1 comprises habit persistence in consumption, DGP 2 relies on GHH preferences with non-time-seperable consumption and hours worked decisions. It can be easily seen from a comparison of Figures 3 and 7 that this leads to qualitatively rather different dynamics of hours worked and consumption in the presence of technological news shocks. The hours

---

[9]Note that we provide graphs in the following discussion only to a limited extent for the sake of confining the number of pages to a convenient number, but the graphs from the alternative simulations are available from the authors upon request.

worked response differs substantially from consumption response under DGP 1, while hours and consumption responses are rather similar under DGP 2. Consumption response to a positive news shock is smooth under DGP 1, whereas there is a sudden and sharp increase in consumption with the realization of the news under DGP 2. On the other hand, the dynamics of TFP is identical across the two DGPs by construction and the dynamics of output rather similar.

In addition to the qualitative differences in the impulse response dynamics, time series generated by DGP 2 exhibit much more variability than time series generated by DGP 1. In our first set of alternative simulations, we bring DGP 2 closer to DGP 1 in terms of its calibration. Thereby, we try out three variations: (1) DGP1 calibration: We set several common parameters to the same value as in DGP 1: $\alpha = 0.33$, $\beta = 0.99$, $\delta''(u) u/\delta'(u) = 0.5549$ and $\phi''(g_A) = 0.3$. Note that the foregoing value of the elasticity of $\delta'(u)$ implies a steady state depreciation rate of 0.025 as in DGP 1, whereas the implied depreciation rate with the benchmark calibration of DGP 2 is much higher than this value. (2) KPR calibration: GHH preferences become identical to KPR preferences in DGP 2 by setting $\gamma = 1$. We try out the impact of this crucial change by leaving all remaining parameters as in the benchmark DGP 2 calibration. (3) DGP1-KPR calibration: Both changes from the DGP1 and KPR calibration are imposed on DGP 2 simultaneously. This last calibration is the one that brings DGP 2 closest to DGP 1 except that there is no habit formation in consumption and variable capital utilization in DGP 1.

DGP 2 indeed generates weaker responses under both DGP1 and KPR calibrations, the responses being weaker under the KPR calibration than the DGP1 calibration. We discuss only the median IRFs from the first and last of these alternative specifications (i.e. DGP1 and DGP1-KPR calibrations) in the following, since the differences between the KPR and DGP1-KPR calibration results are negligibly small.

The impact of the DGP1 calibration on DGP 2 results can be seen by comparing the median IRFs of the variables in Figure 5 with the upper panel of Figure 9. Although the TFP responses are identical in both figures and the responses of consumption, hours worked and output do not change qualitatively, the latter variables exhibit a less pronounced response to news shocks with the DGP1 calibration. Four-variable models are superior to the trivariate models in estimating the median response and BS models slightly outperform BP models with DGP1 calibration too. The most conspicuous observation is, however, that the performance of the VAR models deteriorates when the DGP1 calibration is applied to DGP 2. The relative distances between the true and median responses get wider under the DGP1 calibration.

The DGP1 calibration thus effects both the true dynamics of the variables and the performance of the SVAR models that try to estimate the impact of news shocks on the variables. As to the true dynamics of the variables, their responses to news shocks get further weaker when the DGP1-KPR calibration is imposed on DGP 2 as the graphs in the lower panel of Figure 9 suggest. Moreover, since KPR and DGP1-KPR calibrations lead to similar results, we conclude that it is the KPR specification which leads to the strong weakening of the impulse responses in comparison to the benchmark DGP 2 in the lower panel of Figure 9. In terms of the performance of the SVAR models, we observe that the performances of the three- and four-variable systems generally get very similar under the DGP1-KPR calibration, the four-variable BP model performing slightly worse than the other three VARs, particularly at shorter horizons. In general, all VARs achieve to estimate the IRFs well in both qualitative and quantitative firms.

To sum up, the difference between DGP 1 and DGP 2 with the most crucial effect on the true and estimated dynamics of the variables is the form of the household utility function rather than the parametrization chosen for the DGPs. Particularly, the form of the utility function affects the IRFs of consumption, hours worked and stock price significantly. With GHH preferences as in benchmark DGP 2, four-variable VARs always outperform trivariate models, regardless of whether stock price or consumption is the second variable in the VAR. With KPR preferences, however, the difference between trivariate and four-variable models diminishes substantially. Moreover, switching from GHH to KPR preferences or the preferences in DGP 1 clearly diminishes the size of the IRFs and render SVAR identification easier.

4.2.2   The Optimization Horizon in the BS Scheme

In order to investigate the impact of the length of the optimization horizon on the BS scheme, we try out three alternatives to our benchmark $H = 40$; $H = 20$, $H = 80$ and $H = 200$. In terms of median impulse responses, we find that the optimization horizon does not play a significant role qualitatively, i.e. the shapes of IRFs are not sensitive to this parameter as Beaudry, Nam, and Wang (2011) point out. Quantitatively, however, we obtain much less difference in terms of IRFs with both trivariate and four-variable frameworks, which is in contrast to the findings of Beaudry et al. who report similar results for large H (i.e., results are very similar for $H = 120$ and $H = 200$) that are quantitatively somewhat different from the ones under $H = 40$, for example. We observe only for the trivariate model under DGP 2 and with consumption as the second variable some significant quantitative differences across

the median responses of the variables with respect to different H parameters. Recall that this is at the same time the experiment, where both BP and BS schemes fail substantially in terms of generating good-looking median responses and convenient correlation coefficients between the true and estimated news shocks.

4.2.3 Issues Related to SVARs in General

Sample Size Our simulations highlight once again how important a sufficient sample size is for the success of SVAR identification schemes. We observe a significant deterioration of the SVAR estimates when the sample size is reduced to 100 observations. The median responses illustrated for DGP 1 and empirical models with consumption in the upper panel of Figure 10 are illuminating in this context. There are large discrepancies between true and median responses, particularly at horizons beyond the news shock lead. Furthermore, the VARs generate much wider confidence intervals and obtain much lower correlations between true and estimated news shocks. As to the correlations, they are estimated to be even negative sometimes, whereas with 200 observations in the benchmark simulations all estimated correlations were in the positive domain under DGP 1. A similar deterioration as for DGP 1 is observed for DGP 2 with 100 observations as well.

Increasing the sample size to 1000, on the other hand, leads to striking improvements in the identification performance of the VARs. According to the IRFs computed with VARs with stock price as the second variable and under DGP 1 in the lower panel of Figure 10, for example, the median responses from the four-variable models are very close to the corresponding true responses implied by the DGP. This picture is accompanied by much lower confidence intervals and correlations between true and estimated shocks that lie above 0.8 in all of the simulations we carry out with both DGPs. Nevertheless, a note of caution is in order at this point. A large number of observations alone is not always a recipe for successful VAR identification; the choice of variables entering the VAR must also be correct. The trivariate models, for example, in the lower panel of Figure 10 still substantially underestimate the TFP and hours worked responses. Finally, the differences between BP and BS schemes become negligible with a large sample size regardless of whether they rely on trivariate or four-variable models. The main reason for the relative improvement of the BP schemes is that the matrix of long-run multipliers $\Gamma(1)$ in equations (4) and (5) can be estimated much more successfully than before with a large sample size.

Lag Order Our alternative simulations discussed in the following confirm that the choice of the lag order cannot be independent of the sample size and the properties of the true DGP. Therefore, it is important to recall that our benchmark samples contain 200 observations and the lag order is set to 4 in the benchmark VARs. Remember also from our discussion above that DGP 2 generates much more dispersion in the variables than DGP 1 in the presence of news shocks.

We start out our evaluation with data generated by DGP 1 and estimation carried out with a VAR lag order of 2. Under the relatively smoother setting of DGP 1, the performances of both trivariate models worsen after a reduction of the lag order to 2 in terms of their median IRFs and confidence intervals regardless of whether stock price or consumption enters the VAR. As to the correlation between true and estimated news shocks, however, a lower lag order would lead to better estimates of news shocks with trivariate models.[10] With both four-variable models, on the other hand, the benchmark lag order of 2 generally leads to better results in terms of both IRFs and estimated news shock series.[11] Finally, the IRF and news shock series estimates of BP and BS models with four-variables become virtually identical when the lag order of the underlying VAR is set to 2. This relative performance improvement of the BP scheme might suggest that the estimation of $\Gamma(1)$ gets more successful when the lag order is decreased with a smaller sample size.

When the lag order is 2 but the data are generated under the more volatile setting of DGP 2, we obtain that whether stock price or consumption is the second variable in the VARs plays a much more critical role for the relative success of the VARs vis-à-vis the corresponding benchmark estimates. Since the trivariate models do generally not perform well under DGP 2, we focus only on the four-variable models in our following discussion. When stock price is the second variable in the VAR, BS and BP schemes both generate slightly better (worse) median estimates in the short (long) run. Moreover, four-variable models deliver not narrower but more balanced confidence intervals around the true response when the underlying VAR is estimated with four lags. With consumption as the second variable of the VAR, on the other hand, the median and confidence interval estimates get somewhat better when the lag order is decreased to 2. Finally, the news shock series estimates of the four-variable BP (BS) model improves (worsens) a bit with the lag order of 2 in comparison to the benchmark estimates.

---

[10]The picture a bit mixed for the trivariate BP model though. The share of correlations between true and estimated news shocks above 0.8 decrease, between 0.6 and 0.8 decrease, and less than 0.6 increase.

[11]In the case of four-variable BS model, the improvement in the estimation of news shock series is rather negligible.

When the lag order is set to 6 in the VAR, four-variable models always perform slightly worse than with 4 lags in terms of all measures we consider in this paper and regardless of which DGP is used for data generation. On the other hand, the performances of the trivariate models improve in terms of median IRFs and their confidence intervals when the data is generated by DGP 1 (DGP 2) and the second variable of the VAR is consumption (stock price); in case the data are generated by DGP 1 and consumption is used the VAR, the performances of the trivariate models are not affected from increasing the lag order to 6. Strikingly, the performances of the trivariate models improve significantly and approach to the performance of the four-variable models when DGP 2 is the data generator and stock price is used in the VAR. When the data are generated by DGP 1, the correlations between the true and estimated news shocks generally get somewhat lower for all trivariate models but the BS model with stock price; with data generated by DGP 2 and stock price in the VAR, however, a lag order of 6 leads to higher correlations between true and estimated news shocks.

In order to gain more insight into the impact of sample size and lag order on the performance of BS and BP schemes, we finally use a much longer sample with 1000 observations and increase the lag order to 6. A higher lag order should guarantee that the degrees-of-freedom problem is much less relevant. Interestingly, the performances of the four-variable models either worsen or do not change when we use 6 instead of 4 lags with a sample size of 1000 observations.[12] On the other hand, trivariate models generally show a better identification performance. This suggests that the so-called lag truncation bias is negligible when enough information is contained in the VAR by including more variables, whereas it is a more serious problem for VARs not having enough variables, i.e. enough information.

Barsky and Sims (2011) and Sims (2012) advocate using additional variables, particularly forward-looking ones in VARs in order to ameliorate potential non-invertibility issues. The lesson from our simulations with alternative sample size and lag orders is, however, that including more than four variables in the VARs can lead to a wide dispersion of results such that statistical inference becomes impossible. Rather than estimating a VAR with a large number of variables, estimating VARs with a few variables but including among them ones with particularly strong forward-looking properties seems to be important for successful identification.

---

[12]The only criteria that slightly improves for the four-variable models when 6 lags are used is the correlation between true and estimated news shocks when DGP 2 is the data generator and stock price is used in the VAR.

Cointegration   In our benchmark experiments, we set the cointegration rank to $K-1$ in K-variate VARs by hand in line with our DGPs and estimate the $K-1$ cointegrating vectors with the conventional maximum likelihood (ML) method. In practice, various tests and methods exist to determine the cointegration rank and estimate the cointegrating vectors. While we do not provide an extensive discussion of the issue here, we check to what extent our results change when the cointegrating vectors are changed by their true counterparts as they are known from the DGPs in the VAR estimations. For both DGPs, using the true cointegrating vectors instead of estimated ones leads to a slight improvement of the estimation performance of the VARs such that the median responses get closer to the true ones, confidence intervals get narrower and the average correlation between the true and estimated news shocks increases. Yet, the differences are not big so that practitioners can confidently rely on ML estimates of the cointegrating vectors.

Fisher (2009) notes that the cointegration rank in the model can play an important role for the shape of the impulse response functions. Therefore, Barsky and Sims (2011) prefer to estimate the VAR in levels which produces consistent estimates of the IRFs without the necessity of setting a cointegration rank for the system, which is also the conservative approach suggested by Hamilton (1994). Our findings suggest that the level estimation leads to a (partly large) downward bias in the estimated IRFs, compared to the estimations by means of a cointegrated VAR with a cointegration rank of $K-1$, particularly at horizons beyond the anticipation lead of the news shock. In case the users of SVARs are sure about the existence of cointegrating relationships but not the exact cointegration rank, they might estimate a VECM with a cointegration rank of $K-1$ rather than one in levels, which might lead to a lower bias. It is, however, not possible to investigate the issue further in our framework, since the structure of our DGPs do not allow building SVARs with a cointegration rank less than $K-1$ by construction.[13]

Relative Size of Shocks   Recall that we set the standard deviations of surprise technology (σ_{ε_1}) and technological news shocks (σ_{ε_2}) to 0.66 and the government spending share shocks (σ_{ε_3} and σ_{ε_4}) to 0.25. As **?** show, the relative size of technological news and surprise technology shocks is an important determinant of the performance SVAR identification schemes concerning the detection of these two shocks and their dynamic effects from data.

We reduce the standard deviation of the news shock to 0.33, i.e. the relative size of the

---

[13]All growing variables of the DGPs share the same stochastic trend and all remaining variables are stationary by construction. Any combination of them would imply either a cointegration rank of $K-1$, if the VAR includes at least two growing variables, or an estimation in level, if the VAR includes only stationary variables such as hours worked, the return on capital or capacity utilization.

shock to $\sigma_{\varepsilon_2}/\sigma_{\varepsilon_1} = 0.5$ in our first alternative experiment on the issue. This change either hardly affects the performance of the VARs in terms of median IRFs or leads to a slight qualitative improvement of them, while the confidence intervals of the IRFs do not get wider but more balanced around the true response. However, the VARs' ability for obtaining a high correlation between the true and estimated news shocks deteriorates significantly with a relatively lower size of technological news shocks.

In contrast, when we increase the relative standard deviation of the news shock to 2 by halving the standard deviation of the surprise technology shock to 0.33, we get the opposite picture. We observe that the median IRFs of the trivariate models get slightly better, while the four-variable models' results do not change significantly in this respect. The IRF confidence intervals get, on the other hand, narrower but slightly more asymmetric around the true responses. The most positive impact of increasing the relative size of news shocks is, however, that the news shock series estimates of VARs show a higher correlation with the true ones on average.

? find that the sequential identification scheme of Beaudry and Portier (2006) obtains a higher correlation between news shocks seperately identified with short-run and long-run restrictions when the relative size of news shocks is large. Our finding here resembles the finding of ?: the larger the relative size of news shocks, the higher the correlation between true and estimated news shocks. Our additional insight is that this success of identification schemes comes at the cost of generating less symmetric confidence intervals around the true responses.

Finally, keeping the size of the shocks hitting the technology process as in the benchmark simulation, i.e. a relative standard deviation of unity, and increasing the standard deviation of both government spending share shocks from 0.25 to 0.66 leads to controversial observations. The median IRF estimates of the models get slightly better (worse) when DGP 1 (DGP 2) is the underlying data generating process. This implies that larger government spending shocks makes isolating them from news shocks easier in the relatively more tranquil DGP 1 environment, while they generate more noise in the more volatile environment of DGP 2 and increase estimation uncertainty. Accordingly, whereas the IRF confidence intervals get wider under both DGPs, they get more symmetric (asymmetric) around the true response under DGP 1 (DGP 2). The estimates of news shocks series worsen tough under both GDPs when government spending gets more noisy.

4.2.4   Issues Related to SVARs with News Shock Identification

Anticipation Lead   In the benchmark simulations, the anticipation lead was set to 2 quarters. When the lead is increased to 4 quarters as we do in this subsection, the VARs can still perform rather well in the sense that median IRFs are qualitatively similar to the corresponding true IRFs but, not surprisingly, the quantitative performance deteriorates somewhat. The deterioration in comparison to the benchmark is more conspicuous for trivariate models than four-variable models. This suggests that more information is needed for a good identification of technological news shock dynamics with a longer anticipation lead. The upper panel of Figure 11 shows the median IRFs versus the true ones when the DGP 1 is the data generator and VARs are estimated with consumption as the second variable.

While the confidence intervals generally get more asymmetric, the insights are more diverse concerning the estimation of the technological news shock series, when the anticipation lead is increased from 2 to 4 quarters. The estimates of the series improve under DGP 1, as the lower panel of Figure 11, which shows the histogram of correlations between true and estimated news shock series with DGP 1 and VARs including consumption as the second variable, suggests. The correlations deteriorate, however, when the data are generated by DGP 2, the deterioration rendering the trivariate models useless for econometricians.

A final note of caution relates to the findings of **?** who point out that the lag order of the VAR must be sufficiently high, i.e. at least as high as the anticipation lead of the news shock. In case the lag order is lower than the anticipation lead, the SVAR performance deteriorates substantially.

Measurement of TFP   We address potential problems related to the measurement of the TFP in two alternative ways. First, *we add an AR(1) noise term to the TFP growth rate* which is fairly large with a persistency parameter of 0.5 and due to being hit by exogenous shocks that have half the standard deviation of the surprise and anticipated technology shocks in our framework. This should provide a proxy for addressing the measurement problems related to the TFP being an unobserved variable. Not surprisingly, all benchmark results deterioriate with such a strong noise term. Nevertheless, the SVARs get even more successful in the estimation of IRFs qualitatively, although the distances between true and median responses get wider and the confidence intervals get larger as well as more asymmetric, when the TFP growth rate measurement is subject to noise. Finally, a comparison of the four-variable models under both DGPs shows that the BS scheme gets the median IRF of TFP better, while the BP scheme outperforms the BS scheme in the estimation of the IRFs of

the remaining variables.

Since the availability of TFP, particularly at quarterly frequency, is rather limited in many countries, *using labor productivity as a proxy for TFP* might be a convenient solution.[14] When the TFP is substituted with labour productivity in the VARs, the median IRFs across the BS and BP schemes become almost identical under DGP 1 and the distance between the median IRFs from the BS and BP schemes get much closer under DGP 2. Moreover, the models get more successful than before in getting the IRF patterns of consumption, hours worked and output qualitatively. The problem with proxying the TFP with labour productivity is, however, that both the median IRFs and the confidence intervals show a higher downward-biasedness than before such that the confidence intervals mostly fail to include the true response with the exception of the VARs with consumption as the second variable under the DGP 1. In terms of the estimation of the news shock time series, the picture is mixed: using labour productivity might sometimes lead to a better estimation of the structural shock series.

Recall that the trivariate SVARs perform badly under our benchmark simulation with DGP 2 and where consumption is the second variable of the VAR. A comparison of the median IRFs in the upper panels of Figures 7 and 5 with the upper and lower panels of Figure 12, respectively, shows that using labour productivity instead of TFP may lead to a significant improvements in the qualitative patterns of the SVAR estimates. The bottom line of the discussion for the applied econometrician is that substituting TFP with labor productivity can be a convenient alternative with certain DGPs, but caution is needed in the interpretation of the results since the IRF estimates can be significantly downward-biased with SVARs using labor productivity.

5   Concluding Remarks

The aim of this paper is to investigate how the performances of two alternative SVAR schemes proposed by Beaudry and Portier (2006) and Barsky and Sims (2011) differ in terms of the identification of news shock dynamics. These two studies explore the role played by news shocks in business cycle fluctuations and arrive at controversial conclusions. While the findings of Beaudry and Portier (2006) attribute a significant role to news shocks in business

---

[14]The use of labour productivity in the SVAR literature is quite common. The controversy in the business cycle literature on the role of technology shocks in the business cycle, which started with the paper of Galí (1999), builds on SVARs where labor productivity is the variable that is supposed to account for long run movements in productivity.

cycle fluctuations, Barsky and Sims (2011) obtain opposite results, although both studies find that news shocks are an important driver of long-run trends in the macroeconomic data.

We carry out Monte Carlo exercises with two alternative DGPs in order to address the impact of the nature of the DGP on SVAR identification. The endogenous variables of DGP 1 are much less responsive to technological news shocks of the same size than the variables of DGP 2, and the qualitative patterns of the IRFs are quite different across the two DGPs. Our main finding is that the Beaudry-Portier and Barsky-Sims schemes give quite similar results when the same reduced-form VAR underlies their estimates. With respect to both DGPs, we use two alternative four-variable VARs of which first (TFP), third (hours worked) and fourth (output) variables are identical but the second variable is consumption in one case and stock price in the alternative case. Alternatively, we reduce the VAR order to three variables where output is discarded from the VAR and the second variable of the system is either consumption or stock price.

We find that the SVAR performance is (sometimes heavily) dependent on what type of information is included in the model. This issue and that the BP and BS schemes give similar results under the same information content has recently been addressed by Beaudry and Portier (2013) by using the US data. They point out that estimates of the four-variable model of Barsky and Sims (2011), which comprises TFP, output, consumption and hours worked change substantially, when the model consists of TFP, stock price, output and consumption (or hours worked), and the outcome resembles very much the findings of Beaudry and Portier (2006). Our results can thus be seen as complementary to the insights of Beaudry and Portier (2013).

Barsky and Sims (2011) rightly point out the difficulties of SVAR identification with long-run restrictions by referring to the study of Faust and Leeper (1997), since the coefficients of the matrix of long-run multipliers are then biased in small samples. The critique applies to the the BP identification scheme with three or four variables, since it involves the matrix of long-run multipliers as well. However, like the issue of non-invertibility, the level of bias depends significantly on the DGP and the specification of the VAR as our Monte Carlo simulations demonstrate, and the critique does not require an either/or decision for the users of the BP scheme. In our benchmark simulations, the trivariate and four-variable BS schemes outperform their BP counterparts under both DGPs, particularly at the horizon of roughly up to 16 quarters, while the IRFs from both schemes converge at longer horizons. The difference between the IRFs of the two schemes up to 16 quarters is visually perceivable, yet not so big that it would change the conclusions from the BP and BS schemes.

It turns out that the foregoing quantitative difference between the BP and BS schemes is strongly related to the small-sample bias. When we increase the sample size to 1000 observations, the IRFs from the BP and BS schemes get virtually identical under both DGPs and with both trivariate and four-variable VARs. Moreover, we observe that, increasing the lag order from 4 to 6 with a sample size of 1000, the performance of the four-variable SVARs either slightly worsen or do not change, whereas the performance of the trivariate models clearly improves. These findings highlight the extent to which the curse of dimensionality can reveal itself in VAR models. Even a fairly large sample size of 100 observations appears too short to get reliable results out of VARs with 4 lags, particularly if the true DGP generates a strong response of the endogenous variables to technological news shocks like our DGP 2. This observation leads us to see the findings of Barsky and Sims (2011) with a seven-variable VAR with four lags critically. We note that most of the IRFs reported by Barsky and Sims (2011) from that model with respect to both news and anticipated technology shocks are statistically insignificant.

Another issue that relates to the relative performance of the BP and BS schemes is the measurement of the TFP. We address the issue in two different ways. In the one case, we add a fairly strong noise term to the true TFP growth such that the true TFP series is cointegrated with the TFP series that is subject to noise. Thus, the effects of the noise term would die out over time. The assumption that underlies this approach is that errors in TFP measurement cancel out each other in the long run. In the other case, we use labour productivity as a proxy for TFP, since a TFP series does not exist at quarterly frequency for many countries. With TFP subject to strong noise, the models still perform surprisingly well. With both noise in TFP and labour productivity as a proxy, the estimated IRFs of the VARs get better qualitatively, although their downward-biasedness also increases. Moreover, the distance between the estimates of the BP and BS schemes gets shorter when not the true TFP series enters the VAR. Obviously, the difficulty of measuring the long-run multiplier matrix accurately is then still there for the BP scheme, but the BS scheme is also subject to a similar problem with TFP proxies since the VAR coefficients are estimated less efficiently with TFP proxies than before.

References

Barsky, R. B., and E. R. Sims, 2011, "News Shocks and Business Cycles," Journal of Monetary Economics, 58(3), 273–289.

Beaudry, P., and B. Lucke, 2010, "Letting Different Views about Business Cycles Compete," in NBER Macroeconomics Annual 2009, ed. by D. Acemoglu, K. Rogoff, and M. Woodford, vol. 24, pp. 491–494.

Beaudry, P., D. Nam, and J. Wang, 2011, "Do Mood Swings Drive Business Cycles and Is It Rational?," NBER Working Papers 17651.

Beaudry, P., and F. Portier, 2004, "An Exploration into Pigou's Theory of Cycles," Journal of Monetary Economics, 51(6), 1183–1216.

———, 2006, "Stock Prices, News, and Economic Fluctuations," American Economic Review, 96(4), 1293–1307.

———, 2007, "When Can Changes in Expectations Cause Business Cycle Fluctuations in Neo-Classical Settings?," Journal of Economic Theory, 135(1), 458–477.

———, 2013, "News Driven Business Cycles: Insights and Challenges," NBER Working Papers 19411.

Blanchard, O., J.-P. L'Huillier, and G. Lorenzoni, 2009, "News, Noise, and Fluctuations: An Empirical Exploration," NBER Working Papers 15015.

Chari, V., P. J. Kehoe, and E. R. McGrattan, 2008, "Are Structural VARs with Long-Run Restrictions Useful in Developing Business Cycle Theory?," Journal of Monetary Economics, 55(8), 1337–1352.

Christiano, L. J., M. Eichenbaum, and R. Vigfusson, 2007, "Assessing Structural VARs," in NBER Macroeconomics Annual 2006, ed. by D. Acemoglu, K. Rogoff, and M. Woodford, vol. 21, pp. 1–106.

Cooley, T. F., and M. Dwyer, 1998, "Business Cycle Analysis without Much Theory. A Look at Structural VARs," Journal of Econometrics, 83(1-2), 57–88.

Erceg, C. J., L. Guerrieri, and C. Gust, 2005, "Can Long-Run Restrictions Identify Technology Shocks?," Journal of the European Economic Association, 3(6), 1237–1278.

Faust, J., and E. M. Leeper, 1997, "When Do Long-Run Identifying Restrictions Give Reliable Results?," Journal of Business and Economic Statistics, 15(3), 345–53.

Fève, P., J. Matheron, and J.-G. Sahuc, 2009, "On the Dynamic Implications of News Shocks," Economics Letters, 102(2), 96–98.

Fisher, J. D., 2009, "Comment on 'Letting Different Views about Business Cycles Compete'," in NBER Macroeconomics Annual 2009, ed. by D. Acemoglu, K. Rogoff, and M. Woodford, vol. 24.

Galí, J., 1999, "Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?," American Economic Review, 89(1), 249–271.

Greenwood, J., Z. Hercowitz, and G. W. Huffman, 1988, "Investment, Capacity Utilization, and the Real Business Cycle," American Economic Review, 78(3), 402–17.

Hamilton, J. D., 1994, Time Series Analysis. Princeton University Press.

Jaimovich, N., and S. Rebelo, 2009, "Can News about the Future Drive the Business Cycle?," American Economic Review, 99(4), 1097–1118.

Leeper, E. M., T. B. Walker, and S.-C. S. Yang, 2011, "Foresight and Information Flows," NBER Working Papers 16951.

Schmitt-Grohé, S., and M. Uribe, 2009, "What's News in Business Cycles," CEPR Discussion Papers 7201.

Sims, E., 2012, "News, Non-Invertibility, and Structural VARs," Advances in Econometrics, 28, 81–136.

Sims, E. R., 2009, "Non-Invertibilities and Structural VARs," mimeo, University of Notre Dame.

Figure 1: Response to a News Shock
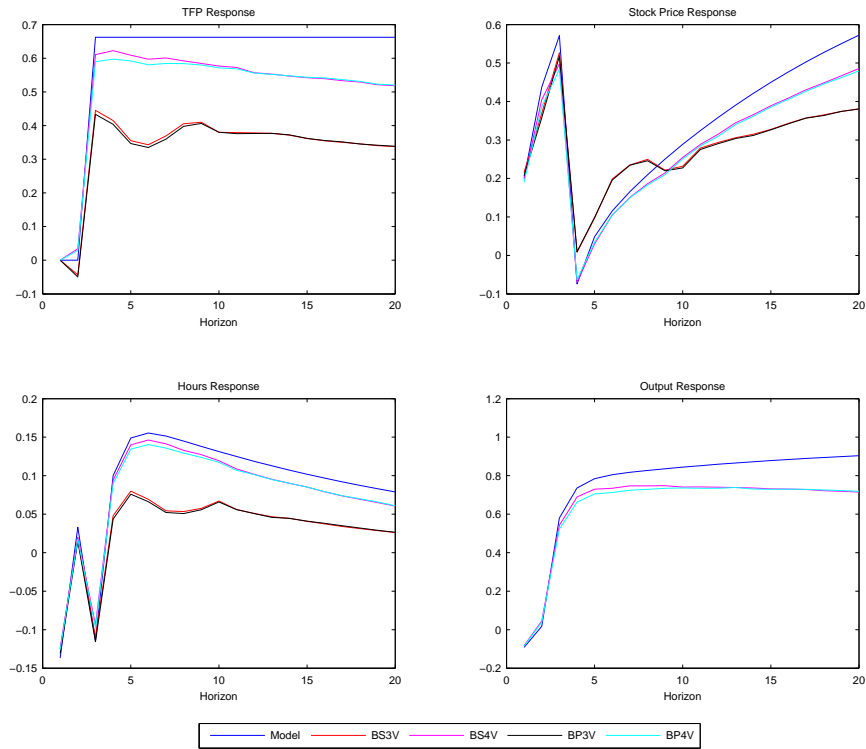
Panel (a): True vs. Median Impulse Responses

Panel (b): True Impulse Response and 90% Confidence Intervals
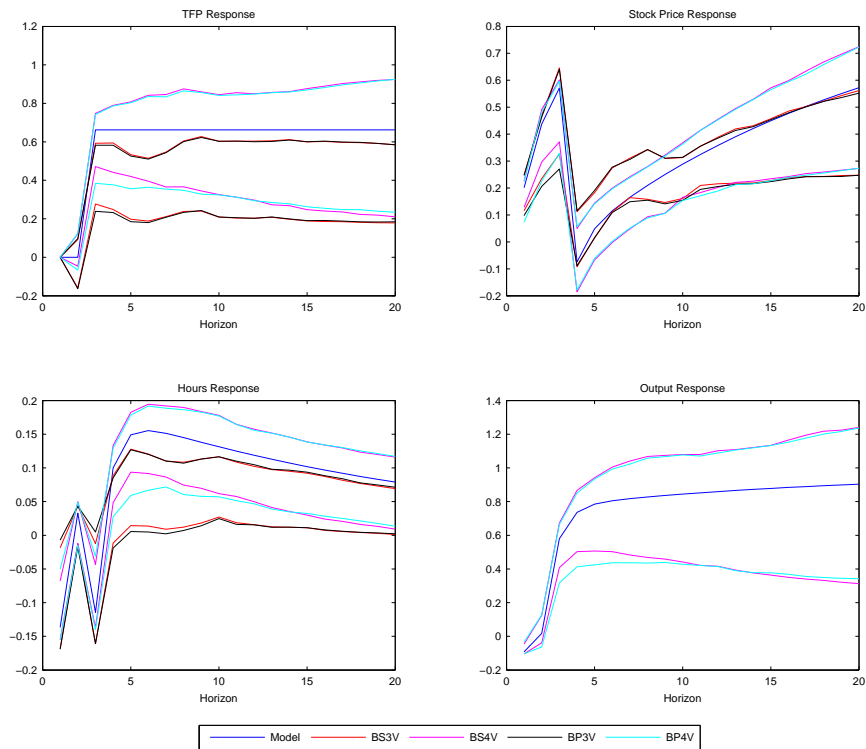
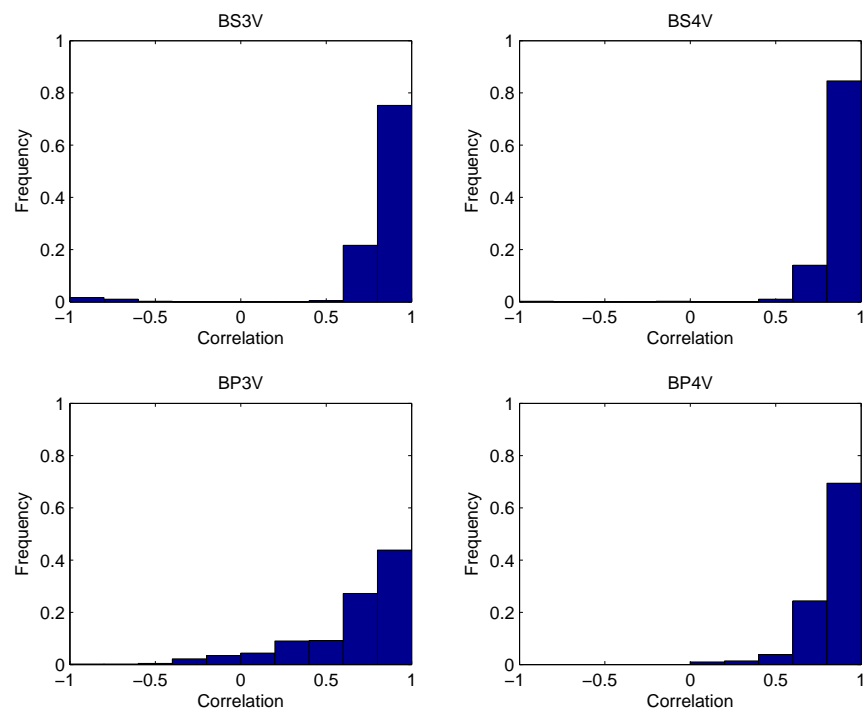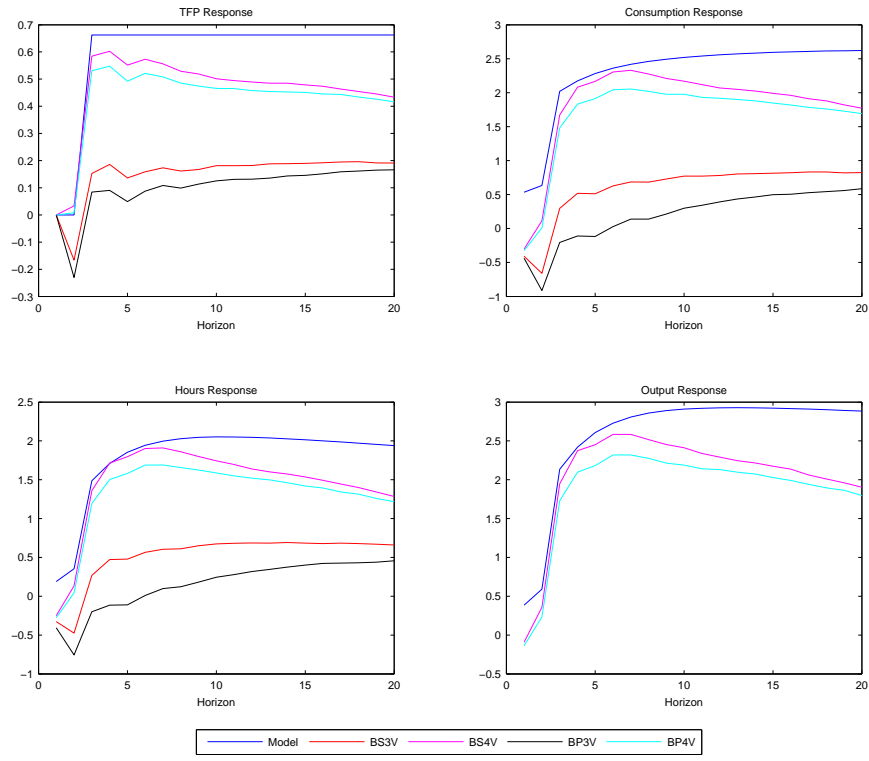*Notes*: Benchmark simulation with DGP 1 where stock price is the second variable. In the upper panel, mean response from 500 simulations is shown for the empirical models. In the lower panel, the 90% confidence intervals are illustrated. The mnemonics "BS" and "BP" respectively refer to Barsky-Sims and Beaudry-Portier models; "xV" denotes the x-variable model.

Figure 2: Distribution of Correlation between True and Empirical News Shocks
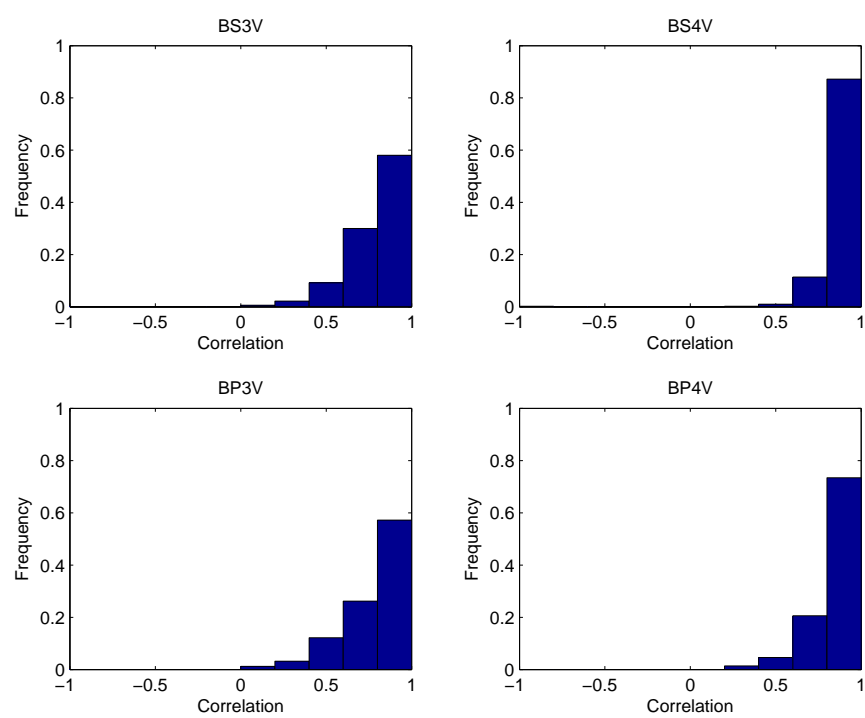
Notes: The probability that the correlation between true and empirical news shocks falls in a certain interval. See Figure 1 for further explanation.

Figure 3: Response to a News Shock
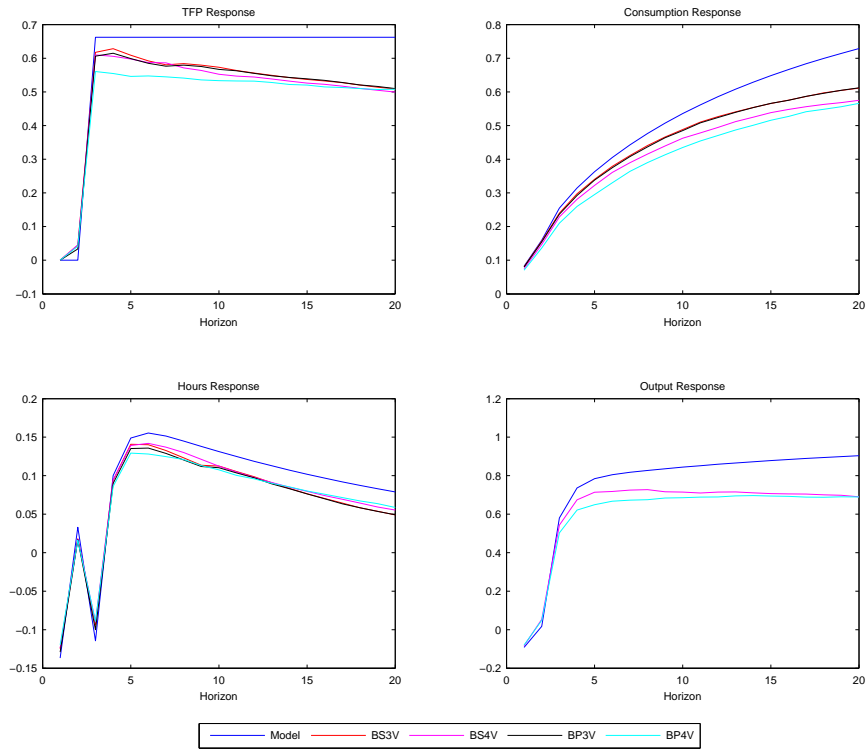
Panel (a): True vs. Median Impulse Responses

Panel (b): True Impulse Response and 90% Confidence Intervals

*Notes*: Benchmark simulation with DGP 1 where consumption is the second variable. See Figure 1 for further explanation.

35
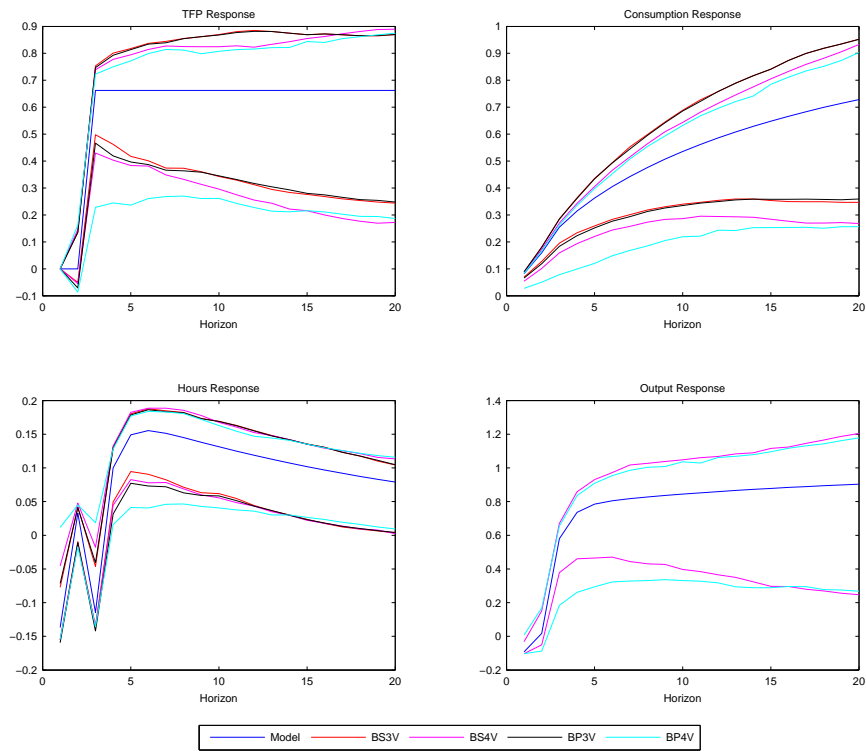
Figure 4: Distribution of Correlation between True and Empirical News Shocks
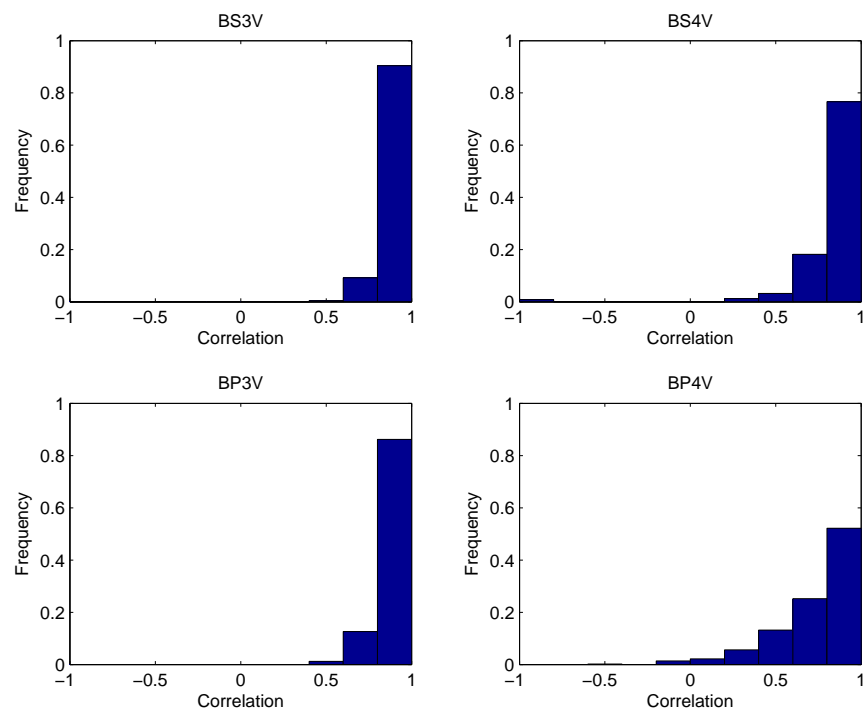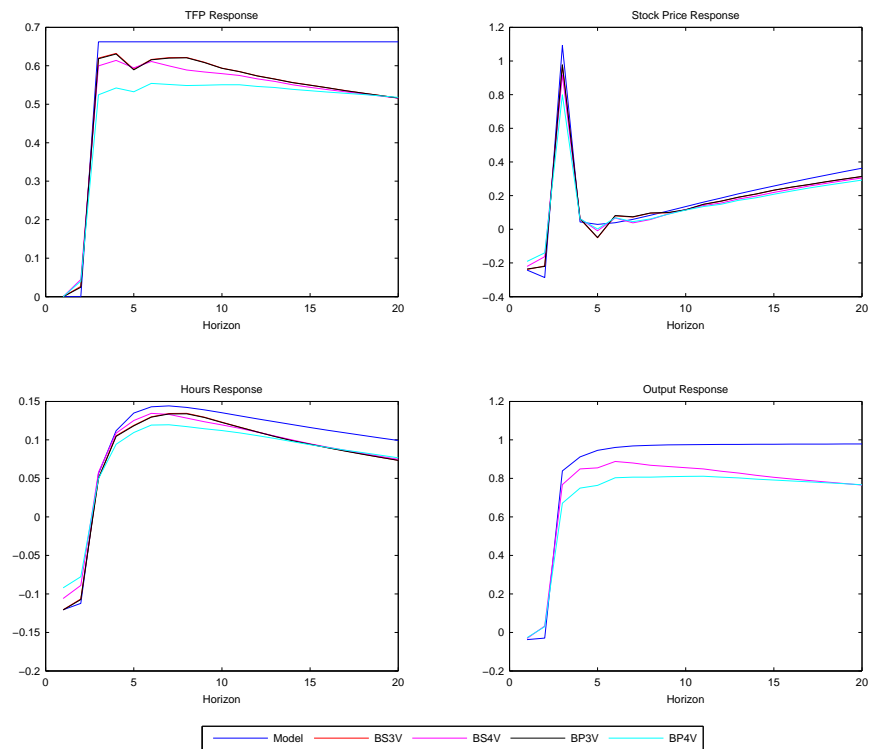

Notes: Benchmark simulation with DGP 1 where consumption is the second variable. See Figure 2 for explanation.

Figure 5: Response to a News Shock
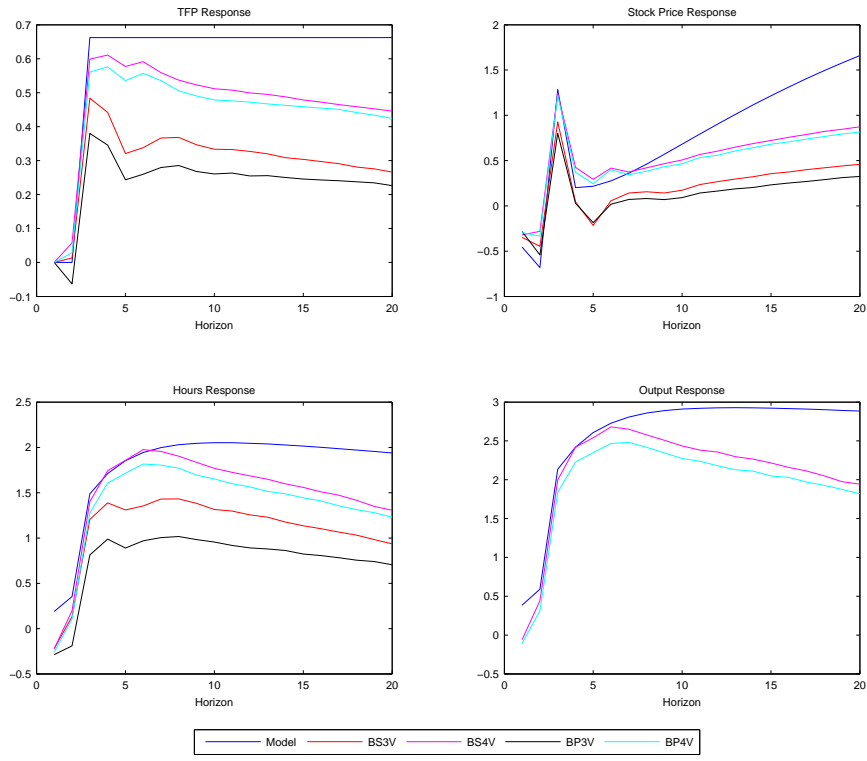
Panel (a): True vs. Median Impulse Responses

Panel (b): True Impulse Response and 90% Confidence Intervals

Notes: Benchmark simulation with DGP 2 where stock price is the second variable. See Figure 1 for further explanation.

37
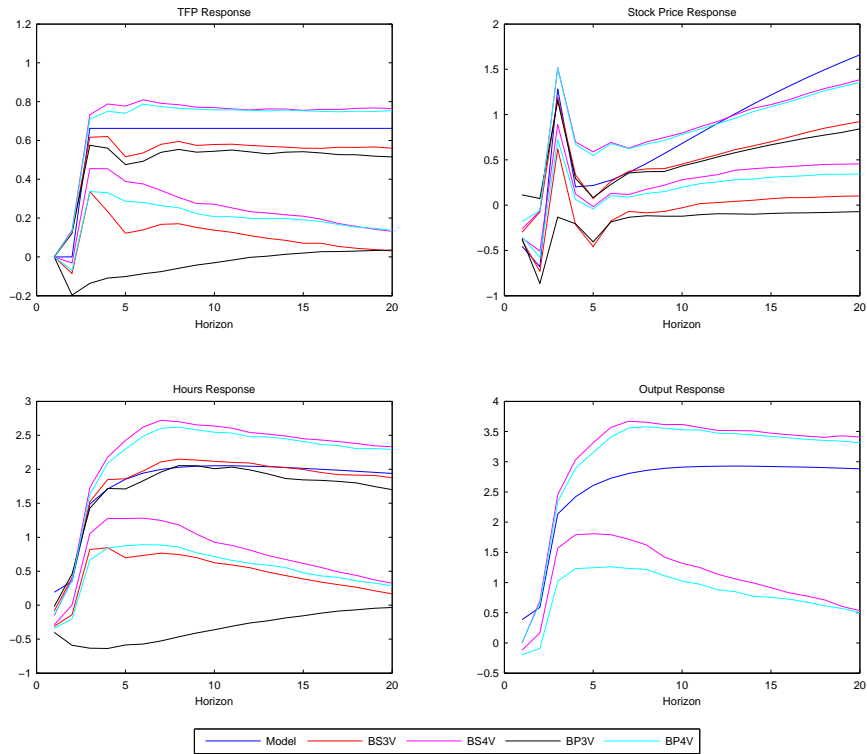
Figure 6: Distribution of Correlation between True and Empirical News Shocks

Notes: Benchmark simulation with DGP 2 where stock price is the second variable. See Figure 2 for explanation.

Figure 7: Response to a News Shock

Panel (a): True vs. Median Impulse Responses

Panel (b): True Impulse Response and 90% Confidence Intervals

Notes: Benchmark simulation with DGP 2 where consumption is the second variable. See Figure 1 for further explanation.

39
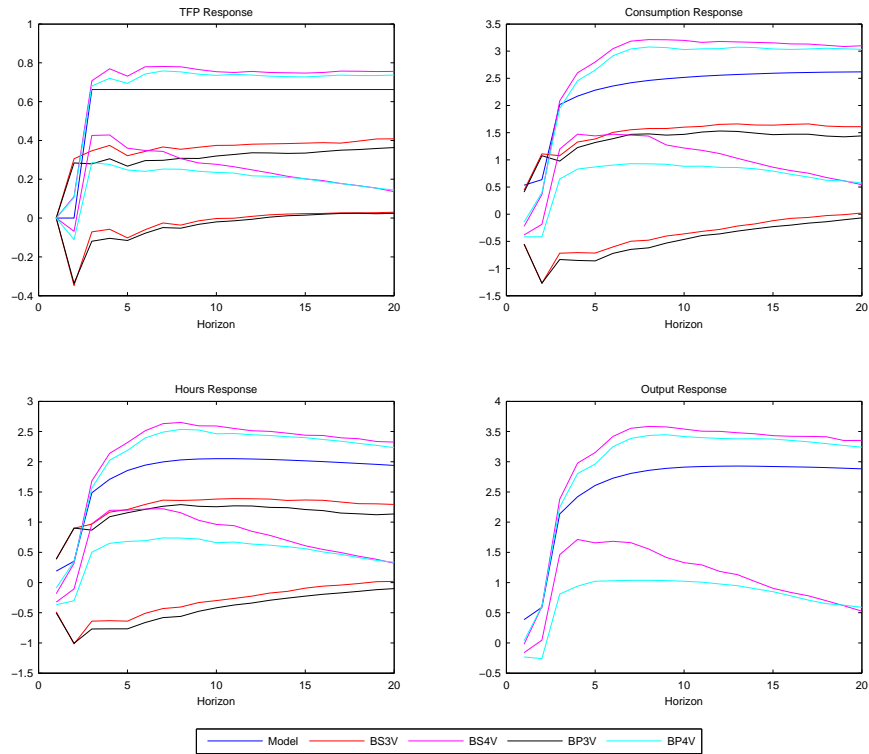
Figure 8: Distribution of Correlation between True and Empirical News Shocks
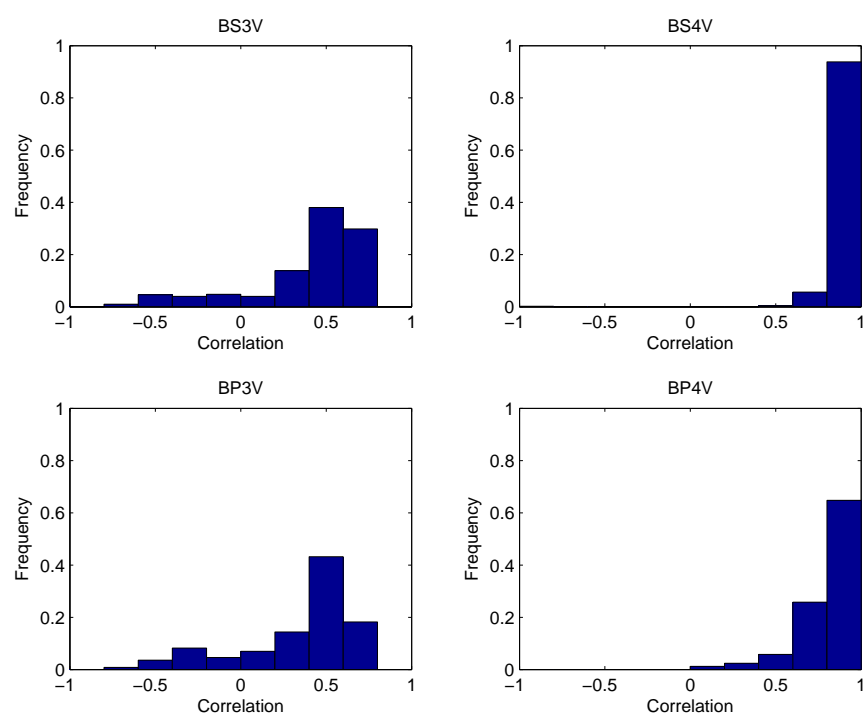
Notes: Benchmark simulation with DGP 2 where consumption is the second variable. See Figure 2 for explanation.

Figure 9: Response to a News Shock

Panel (a): DGP1 Calibration
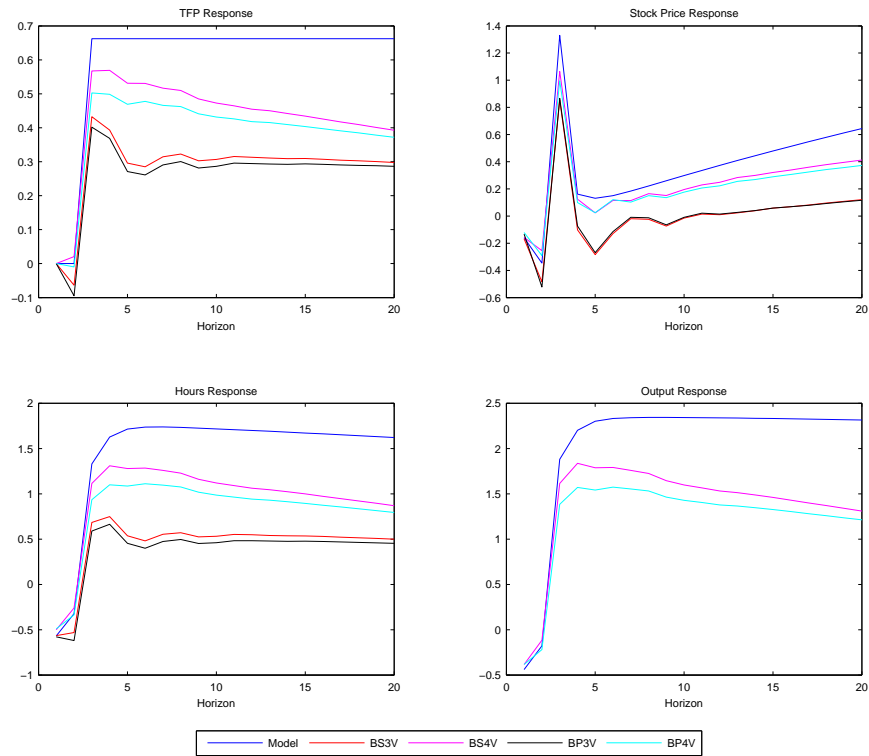
Panel (b): DGP1-KPR Calibration

Notes: Alternative simulation with DGP 2 with DGP1 and DGP1-KPR calibration. The second variable in the VARs is stock price. See Figure 1 for further explanation.

Figure 10: Response to a News Shock

Panel (a): 100 Observations in the Sample
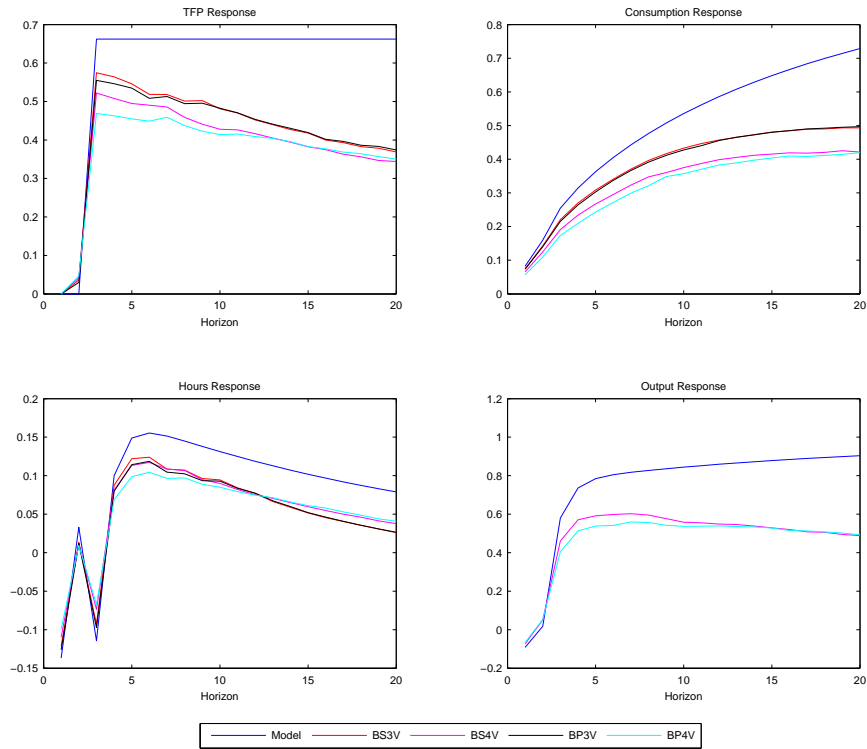

Panel (b): 1000 Observations in the Sample
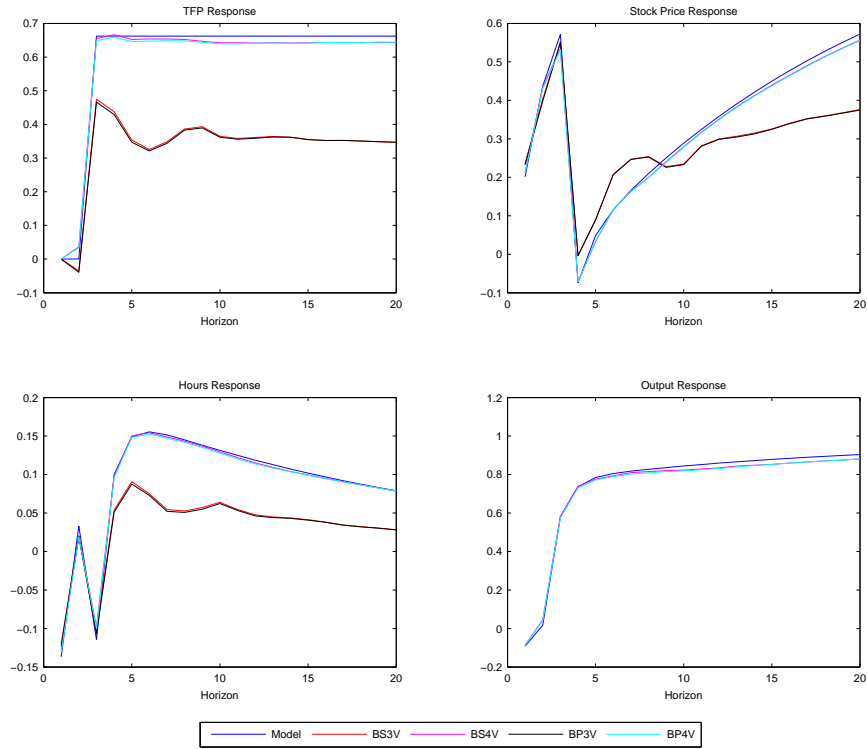
Notes: Alternative simulation with DGP 1 with 100 (1000) observations in the sample. Consumption (stock price) is the second variable in the VAR in the upper (lower) panel. Sample with 100 (1000) observations underlie the estimates in the upper (lower) panel. See Figure 1 for further explanation.

Figure 11: Response to a News Shock
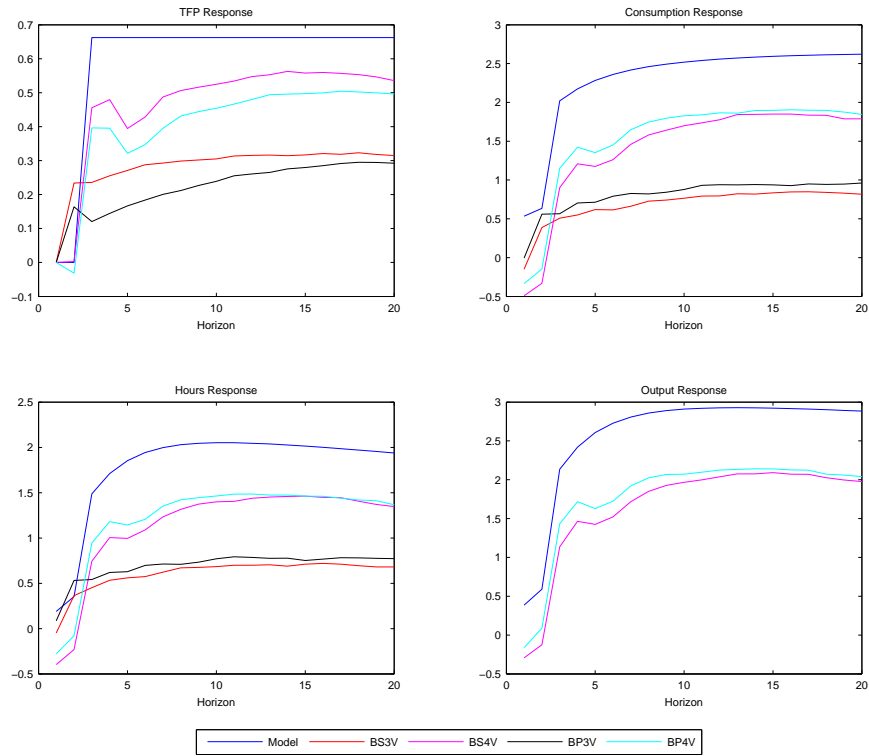
Panel (a): True vs. Median IRFs
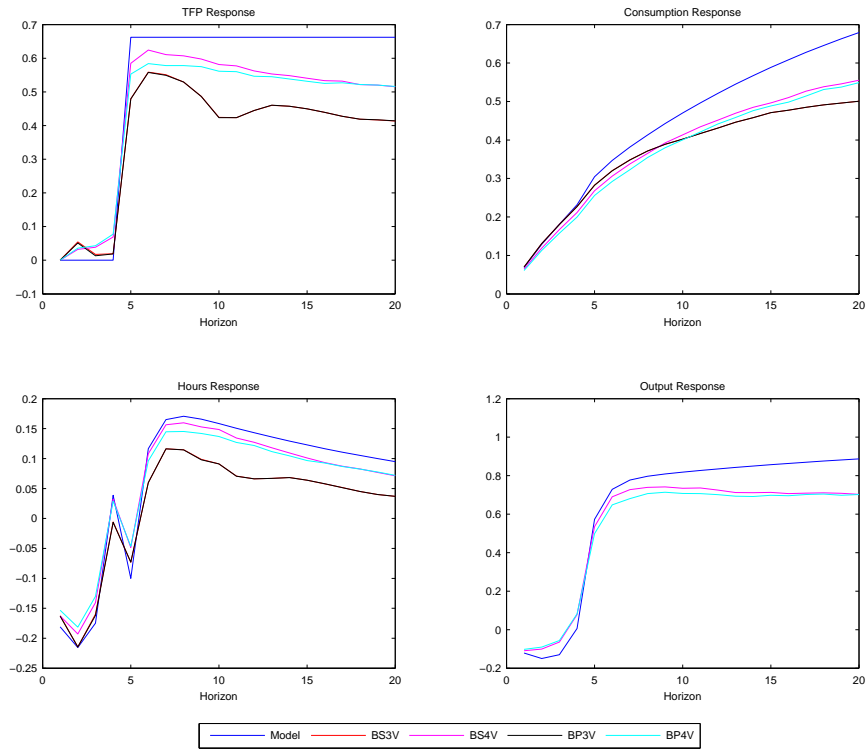
Panel (b): Histogram of Correlations between True and Estimated Shocks
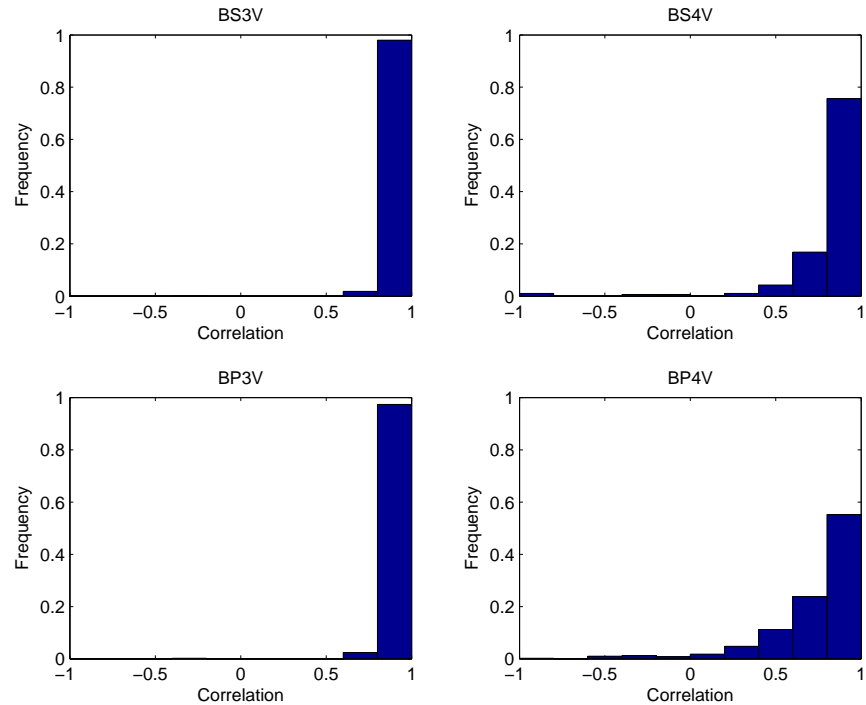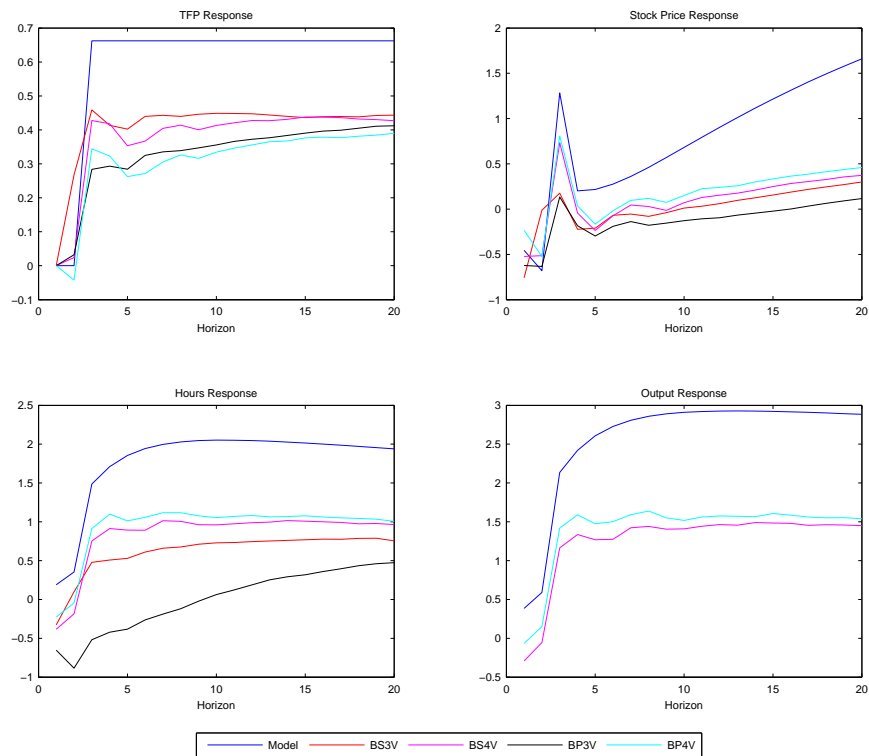

Notes: Alternative simulation with technological news shock anticipation lead increased to 4 quarters. See Figure 1 for further explanation.

43

Figure 12: Response to a News Shock

Panel (a): Consumption the Second Variable in the VAR

Panel (b): Stock Price the Second Variable in the VAR

Notes: Alternative simulation where the measurement of TFP growth rates is subject to noise. See Figure 1 for further explanation.