Does the cost channel matter for inflation dynamics?
An identification robust structural econometric analysis

Rolf Scheufele*

First version: April 2009
This version: November 2010

Abstract

This paper investigates whether interest rate changes impact on firms' marginal costs and whether this has a direct effect on their price setting behavior, translating into aggregate inflation dynamics. Empirical tests on the existence of the cost channel are employed, using a structural econometric approach. Estimation and inference is conducted using identification robust methods based on the continuous-updating GMM objective function. We document identification difficulties for some parameters when estimating the general model structure. For the U.S. a pure forward-looking interest rate augmented Phillips curve is the most compatible with data. This suggests that considering the cost channel is a relevant aspect for monetary policy analysis.

*I would like to thank Joerg Breitung, Katja Drechsel, Jean-Marie Dufour, Sebastian Giesen, Uwe Hassler, Katerina Koka, Dieter Nautz, Chris Sims Jan-Egbert Sturm and Klaus Wohlrabe for many helpful suggestions. I have also benefited from comments of seminar participants at the Macroeconomic Research Meeting (MAREM) 2009 in Tuebingen, the 8th Workshop "Money, Banking, and Financial Markets" in Duesseldorf, the ZEW Macroeconomics Conference 2009 in Mannheim, the 2009 Far East and South Asia Meeting of the Econometric Society (FESAMES 2009) in Tokyo, the 2009 Meeting of the German Economic Association (VfS) in Magdeburg, the 2010 PhD Meeting of the Royal Economic Society in London, the Society for Nonlinear Dynamics and Econometrics 18th Annual Symposium (SNDE2010), the Spring Meeting of Young Economists (SMYE) 2010 in Luxembourg, the 14th Annual International Conference on Macroeconomic Analysis and International Finance (ICMAIF) 2010 in Crete and the EEA annual conference 2010 in Glasgow. Contact details: Halle Institute for Economic Research (IWH), Macroeconomics Department, P.O. Box 11 03 61, D-06017 Halle (Saale), e-mail: Rolf.Scheufele@iwh-halle.de, phone: +49(0) 345 7753-728.

Contents

1 Introduction

This paper investigates whether the costs of external funds affect firms' marginal costs, thereby influencing the aggregate inflation rate. Recently, many authors – including Christiano, Eichenbaum and Evans (2005), Chowdhury, Hoffmann and Schabert (2006), Ravenna and Walsh (2006) and Tillmann (2008) – provide evidence of a cost channel relevant to inflation dynamics. This cost channel is introduced through the cost of working capital into the standard New Keynesian model, which is motivated by cash-in-advance, i.e. factors of production, which have to be paid before the proceeds from the sale of output are received. Empirically, the existence of a cost channel can be tested by augmenting the New Keynesian Phillips curve with an interest rate variable as an additional regressor. So the cost channel implies an extension of the standard measure of marginal cost by interest rate effects.

Chowdhury et al. (2006) test such an augmented Phillips curve specification for G7 countries with GMM and find empirical support for this model for most of them. Ravenna and Walsh (2006) employ the same method, but instead of relying on the reduced form parameters, they estimate the structural parameters of a pure forward-looking specification for the U.S. and draw similar conclusions. The existence of a cost channel is also supported by methods of indirect inference (e.g., Christiano et al., 2005; Huelsewig, Henzel, Wollmershaeuser and Mayer, 2009). However, other studies cast doubt on the existence of a cost channel (e.g., Rabanal, 2007; Gabriel, Levine, Spence and Yang, 2008). Their estimation includes both Bayesian Methods (Rabanal, 2007) and GMM (Gabriel et al., 2008).

In this paper, we extend the Phillips curve specification of Ravenna and Walsh

(2006) to a model that allows for backward-looking behavior in price setting owing to partial indexation. Then we reexamine the existence of the cost channel by estimating structural form parameters for the U.S. (similar to Ravenna and Walsh, 2006). Instead of relying on a standard two-step GMM estimator, we use a continuous-updating GMM (CUE) estimator as proposed by Hansen, Heaton and Yaron (1996). This estimator is preferable in terms of small sample properties, and it does not depend on the normalization of the orthogonality conditions. Moreover, we combine several additional tools in order to analyze the empirical model in great detail. We use identification robust econometric techniques, which can be readily based on the CUE objective function (as suggested by Stock and Wright, 2000). So this procedure guards against problems which are induced by weak instruments and which might be present in estimates of the new Phillips curve (see Ma, 2002; Mavroeidis, 2005; Dufour, Khalaf and Kichian, 2006). Confidence intervals for the individual parameters are then computed by means of the projection technique. We also apply simulation techniques to analyze the complete distribution of the structural parameters. So that the available information of private agents can be used most effectively, factors are used as additional instrumental variables (as proposed by Bai and Ng, 2010; Kapetanios and Marcellino, 2010). To facilitate the comparison of non-nested models, we also make use of Andrews and Lu's (2001) model selection criteria for GMM estimation.

The results of this paper indicate that empirical evidence of the cost channel is not as clear-cut as previously indicated in literature. Generally, the standard procedure of testing one parameter as significant (which is then interpreted as evidence of a cost channel) is found to be inappropriate owing to substantial identification problems. We find that the structure of the model does not allow for drawing strong conclusions about certain aspects of the model on account of weak identification. However, we are able to compare a standard Phillips curve model with the cost channel augmented version, with interesting findings. In fact, the estimated degree of nominal rigidity in a cost channel augmented Phillips curve model is more in line with economically plausible values, and turns out to be statistically more reliable. Additionally, model selection criteria clearly favor a model in which the cost channel is present and where the bank lending rate is included as an additional variable in the marginal cost measure.

Although distinguishing between forward-looking and backward-looking behavior proved difficult, we also document that backward-looking aspects (in the form of indexation) seem to be negligible in all the versions of the model. For the U.S. therefore, a purely forward-looking model that considers the cost channel is most compatible with the data. The inclusion of a cost channel can indeed improve the reliability of estimates of the New Keynesian Phillips curve by introducing the interest rate affecting real marginal costs. Interestingly, this model incorporates enough inflation persistence even without a lagged inflation term (which would in any case be difficult to motivate according to proper microfoundations).

The rest of the paper is organized as follows. Section 2 introduces the theoretical model setup. The empirical strategy is outlined in Section 3. Section 4 presents the estimation results of the interest-rate augmented Phillips curve. Section 5 concludes.

2 The basic model

This section briefly introduces the theoretical model, which consists of a standard New Keynesian framework. More detailed derivations may be found in Walsh (2003) and Woodford (2003). We concentrate on the aspects necessary for characterizing inflation dynamics in the economy. The two basic model features are monopolistically competitive goods markets and sticky prices. In addition, the cost channel is introduced.

More precisely, the economy consists of a continuum of firms (indexed by $i \in [0,1]$), each producing a differentiated good $Y_t(i)$ according to a standard Cobb-Douglas production function

$$Y_t(i) = A_t \overline{K}_t(i)^{\alpha} N_t(i)^{1-\alpha}, \tag{1}$$

with A_t a common country-wide technological factor, $\overline{K}_t(i)$ the (fixed) firm-specific capital stock and $N_t(i)$ denoting the labor factor employed by firm i.

Each firm i faces a demand function characterized by the constant elasticity of substitution given by

$$Y_t(i) = \left(\frac{P_t(i)}{P_t}\right)^{-\epsilon} Y_t, \tag{2}$$

where Y_t equals the aggregate demand, P_t is the aggregate price level in the economy and $P_t(i)$ is the price of good i charged by firm i. The price elasticity of demand for good i is characterized by the parameter ϵ (with $\epsilon > 1$). This determines the constant mark-up (defined as $\mu = \epsilon/(\epsilon - 1)$) required by firms over nominal marginal costs of inputs.

Next, we introduce a liquidity constraint for firms operating in their factor markets. Input factors like the wage bill have to be paid before revenues for the goods produced have been received. To meet these expenditures, firms have to borrow the outlays from a financial intermediary sector. In each period, the individual firm i is assumed to borrow the amount of $Z_t(i)$ to repay the sum total of the salaries. So the liquidity constraint is given by

$$Z_t(i) \geq W_t N_t(i),$$

with W_t the nominal wage rate and $N_t(i)$ the utilized labor factor of firm i. At the end of the period, when the produced good has been sold, firms have to repay these loans with interest to the amount of $i_t^l Z_t(i)$. With these liquidity constraints, firms' marginal costs are equal to

$$MC_t(i) = \frac{R_t^l W_t / P(i)_t}{(1-\alpha) Y_t(i) / N_t(i)} = \frac{R_t^l S_t(i)}{(1-\alpha)}, \tag{3}$$

where $R_t^l = 1 + i_t^l$ and $S_t(i)$ is the firm's specific labor share of production.

Further, we assume that firms face nominal price rigidities which can be characterized by Calvo's (1983) model of staggered price setting. This model implies that firms set prices infrequently owing to the costs of gathering information. The frequency of price re-optimizations is characterized by a stochastic process, with the constant probability that a firm changes its price at one particular point in time. So on the aggregate level at each point in time there is a fraction of firms' $1-\theta$ that optimally adjusts prices. The expected waiting period is then given by $1/(1-\theta)$.

Price re-optimizing firms that set their optimal price $P_t^*(i)$ are faced with the following dynamic maximization problem:

$$E_t \sum_{k=0}^{\infty} (\beta\theta)^k \upsilon_{t,t+k} \left[P_t^*(i) X_{t,t+k} - MC_{t,t+k}(i) \right] \frac{Y_{t+k}(i)}{P_{t+k}}, \tag{4}$$

subject to the demand constraints (eq. 2) and

$$X_{t,t+k} = \begin{cases} \prod_{l=0}^{k-1} \overline{\pi}^{1-\xi} \pi_{t+l}^{\xi} & \text{for } k > 0 \\ 1 & \text{for } k = 0. \end{cases} \tag{5}$$

with β a constant discount factor, $\upsilon_{t,t+k} = U^{'}(C_t)/U^{'}(C_{t+k})$ the time-varying portion of the discount factor between t and $t+k$; with $U^{'}(C_t)$ being the marginal utility of consumption. $\overline{\pi}$ denotes the long-run average gross rate of inflation. Whenever a firm does not re-optimize its price, it is reset according to an indexation scheme. $\xi \in [0,1]$ measures the degree of indexation to past inflation rates. Note that this partial indexation scheme nests more specific indexation assumptions as special cases.[1]

As shown by Walsh (2003), aggregate inflation $\hat{\pi}$ can be related to average real marginal cost $\widehat{mc}$ according to

$$\hat{\pi}_t = \gamma_f E_t \hat{\pi}_{t+1} + \gamma_b \hat{\pi}_{t-1} + \lambda \widehat{mc}_t, \tag{6}$$

where

[1] This specification is adopted from Smets and Wouters's (2003). With $\xi = 1$ it equals Christiano et al.'s (2005) dynamic indexation scheme, with $\xi = 0$ it simplifies to a pure forward-looking model with an indexation to trend inflation.

$$\lambda = \frac{(1-\theta\beta)(1-\theta)}{(1+\beta\xi)\theta}\frac{1-\alpha}{1+\alpha(\epsilon-1)},$$
$$\gamma_f = \frac{\beta}{1+\beta\xi},$$
$$\gamma_b = \frac{\xi}{1+\beta\xi}.$$

This inflation equation is known as the *Hybrid New Keynesian Phillips curve*. Its reduced form coefficients γ_f, γ_b and λ are non-linear functions of the structural parameters β, θ, ξ, α and ϵ.[2] When $\xi = 0$, the equation reduces to the pure forward-looking New Keynesian Phillips curve. When the cost channel is introduced, real marginal costs depend not only on the labor share of output (as derived by Galí and Gertler, 1999) but also on the nominal interest rate:

$$\widehat{mc}_t = \hat{R}^l_t + \hat{s}_t,$$

where $\hat{s}_t = \hat{w}_t + \hat{n}_t - \hat{y}_t$ is the log deviation of the labor share around the steady state and $\hat{R}^l_t$ is the percentage point deviation of the nominal interest rate (defined as the lending rate) around its steady state value.

According to Chowdhury et al. (2006), it is assumed that the lending rate R^l_t can deviate from the nominal interest rate set by monetary policy which is denoted by R^m_t. This is motivated by financial market imperfections and is, for instance, also motivated by the likelihood of defaults. Adopting the simplified framework of Chowdhury et al. (2006), where profit maximization of financial intermediaries

[2]Note that there are also other versions of the structural Phillips curve that have a slightly different interpretation (e.g., Galí and Gertler, 1999). As shown by Scheufele (2010), Galí and Gertler's (1999) model leads to conclusions similar to those considered here.

leads to a log linear relationship between the risk-free rate $\hat{R}^m$ (which is assumed to be under the control of monetary policy) and the lending rate $\hat{R}^l$. This is given by

$$\hat{R}^l_t = (1 + \psi_R)\hat{R}^m_t, \tag{7}$$

where the coefficient $1+\psi_R$ measures the response of the lending rate $\hat{R}^l_t$ to changes in the monetary policy rate $\hat{R}^m_t$. As illustrated by Chowdhury et al. (2006) for $\psi_R > 0$, indicating the existence of strong financial market imperfections. When the opposite holds ($\psi_R < 0$), then management costs are very high.

Now we can express the Phillips curve as a function of the labor share as well as of the monetary policy rate, which is given by

$$\hat{\pi}_t = \gamma_f E_t \hat{\pi}_{t+1} + \gamma_b \hat{\pi}_{t-1} + \lambda \hat{s}_t + \lambda \phi_m \hat{R}^m_t, \tag{8}$$

where $\phi_m = (1 + \psi_R)$. So the idea of the cost channel of monetary transmission follows directly from this equation: whenever the central bank raises its interest rate above its steady state level, this leads to an increase in the current inflation rate over its steady state value. This holds true unless this effect is not overcompensated by the response of the labor share through adjustments of aggregate demand.

3 Empirical analysis

Next we introduce the strategy for estimating the interest-rate augmented Phillips curve specification and how it is possible to conduct inference about the parameters of interest. Generally, the choice is between two different econometric methods: full information or limited information methods. The choice between these categories has a long history in econometrics. Full information methods provide the complete range of statistical properties associated with the model under investigation. This is preferable in terms of efficiency, unless the model is correctly specified. Limited information methods do not require a fully specified model, but it is enough to set up certain moment conditions to estimate the parameters of interest. Thus there is the classical trade-off between efficiency and the sensitivity to model misspecifications known from simultaneous equations models. Since we are interested solely in the Phillips curve equation and more specifically in the direct impact of interest rates on inflation, we find it more convenient to use limited information methods, because we do not want to restrict our results to a particular model structure.[3] Additionally, the great advantage of limited information methods is that identification issues for these types of techniques are now well established. On the other hand, there is no widely accepted method in full information for dealing with problems of weak identification.

3.1 The basic econometric specification

Limited information methods typically require the application of instrumental variable (IV) estimation methods. To get an empirical traceable specification from the theoretical model eq. (8), the unobserved variable $E_t\hat{\pi}_{t+1}$ is replaced by its realization assuming the forecasting error $\eta_{t+1} = [E_t\hat{\pi}_{t+1} - \hat{\pi}_{t+1}]$ to be orthogonal to

[3] Since we also stress the importance of identification robust inference, full information methods like ML are not immune from that kind of problem. However, certain LI methods are able to deal with them. For this reason, once weak identification problems appear, ML, with its asymptotic theory, can generally not be relied on, and there are no full information methods that are identification robust.

past information. So we obtain the estimable equation

$$\hat{\pi}_t = \frac{\beta}{1+\beta\xi}\hat{\pi}_{t+1} + \frac{\xi}{1+\beta\xi}\hat{\pi}_{t-1} + \frac{(1-\theta\beta)(1-\theta)}{(1+\beta\xi)\theta}\left(\hat{s}_t + \phi_m \hat{R}_t^m\right) + u_t, \qquad (9)$$

where $u_t = \nu_t + \gamma_f \eta_{t+1}$. We allow the error term to follow a very general structure - so u_t may be autocorrelated and / or heteroskedastic.[4] The natural setup for estimating potentially non-linear dynamic models is to employ GMM as proposed by Hansen (1982). With the assumption $E_{t-1}u_t = 0$, the moment conditions are given by $E_{t-1}\{f_t(\vartheta)\}$, where $f_t(\vartheta) = u_t(\vartheta)\mathbf{z_{t-1}}$ with $\mathbf{z_{t-1}}$ the vector of instruments including predetermined variables dated $t-1$ or earlier. $\vartheta = (\beta, \theta, \xi, \alpha, \epsilon, \phi^i)$ denotes the parameter vector of interest.

When it comes to parameter estimation, we do not consider the convenient two-step GMM (2GMM) estimator that is frequently used for estimating NKPC models (see e.g., Galí and Gertler, 1999; Galí, Gertler and López-Salido, 2001; Eichenbaum and Fisher, 2007). Instead, we adhere to the continuous updating GMM (CUE) estimator as proposed by Hansen et al. (1996). This estimator is superior in terms of finite sample properties (Hansen et al., 1996; Stock and Wright, 2000). It is more closely related to LIML than to 2SLS (which is related to the 2GMM estimator).[5] Moreover, it does not share the property of standard GMM that estimation bias increases with the inclusion of irrelevant instruments (as documented by Tauchen, 1986; Kocherlakota, 1990). For non-linear settings, another favorable

[4]When we assume ν_t to be white noise and use instruments dated $t-1$ or earlier, then u_t follows an MA(1) process per construction. See appendix A.1 for a discussion.

[5]For the estimation of a single equation in the linear simultaneous equation model, the two-step GMM estimator is 2SLS, whereas the continuous updating estimator is LIML. The superior characteristics of LIML in comparison with 2SLS in finite samples has been well documented in the literature (see e.g., Judge, Griffiths, Hill, Luetkepohl and Lee, 1985, Chapter 15).

property is its insensitivity to the statement of the moment conditions. Since the NKPC in its structural formulation is non-linear in its parameters, it is possible to reformulate the orthogonality conditions, for instance through multiplying by a certain parameter. 2GMM estimates may be sensitive to this kind of transformation (see Hall, 2005, for a general discussion and Scheufele, 2008, for this problem in the context of the NKPC).

The CUE estimates can be obtained by minimizing the objective function

$$\mathbf{S}(\vartheta) = \left[\frac{1}{T}\sum_{t=1}^{T} f_t(\vartheta)\right]' W(\vartheta)^{-1} \left[\frac{1}{T}\sum_{t=1}^{T} f_t(\vartheta)\right], \tag{10}$$

where $W(\vartheta)$ is a $k \times k$ positive semi definite weighting matrix. It can be shown that the weighting matrix is given by the inverse of the asymptotic variance matrix S_T^{-1}. This matrix is computed to be heteroscedastic and autocorrelation consistent (HAC), as proposed by Newey and West (1987). The peculiarity of this estimator is that the covariance is estimated together with the parameter vector ϑ. Instead, the 2GMM computes first an initial estimate of ϑ with a pre-specified weighting matrix (e.g., the identity matrix) and then uses this initial estimate to specify the weighting matrix in the second step.

To estimate the structural form parameters, it is necessary to calibrate some parameters, as when there are four variables one can at least identify the same number of parameters. Like Galí et al. (2001), we choose to calibrate α and ϵ (like them, we set $\alpha = 0.270$ and $\epsilon = 11$ for the U.S. economy). Given these values, the point estimates for β, θ, ξ and ϕ_i can be computed.

3.2 Coping with potentially weak instruments

Once the empirical model has been set up, it must be remembered that the GMM approach is extremely susceptible to problems brought about by weak instruments (see e.g., Stock, Wright and Yogo, 2002; Dufour, 2003; Andrews and Stock, 2005, for recent surveys). These may arise whenever the instruments are not sufficiently correlated with the variables they instrument. This pathology results in parameter distributions that may be far from normality. This leads standard test statistics (e.g., Wald and significance tests) to spurious over-rejections and may lead to wrong conclusions.

Broadly speaking, weak instrument problems do result in identification difficulties for relevant parameters. For example, consider the standard linear instrumental variable (IV) model in the extreme case of an exact zero correlation of the instruments with the right-hand side endogenous variable(s). Obviously, this is a violation of the rank condition for identification (implying that the coefficient in the structural equation is unidentified). However, a correlation of exactly zero is never seen. So problems occur even when the correlation between the instruments and the instrumented variables is not strong enough (then the parameters are close to being unidentified).

It is also important to stress that the problem of weak identification is not specific to GMM or IV estimation. Instead, ML-methods or other methods of Matching Moments may also be affected by this pathology. Because of their high-dimensionality and their non-linear structure, the parameters of DSGE models have generally been found difficult to identify (see e.g., Canova and Sala, 2009, for a general discussion for this aspect).

In order to avoid the problems of weak identification, we combine two empirical strategies. First, we carry out a careful selection of relevant instruments $\mathbf{z}$. Like Bai and Ng (2010) and Kapetanios and Marcellino (2010), we apply Factor-GMM estimation, whereby factors are extracted from a large data set of macroeconomic indicators and used as additional instrumental variables. The advantage here is that instruments can be used parsimoniously, but factors still provide a great deal of informational content. Our second strategy is to rely on identification robust inference methods that are robust to the problems associated with weak instruments. These methods have been successfully applied to the standard NKPC by Ma (2002), Dufour et al. (2006), Mavroeidis (2007), Nason and Smith (2008) Martins and Gabriel (2009) and Kleibergen and Mavroeidis (2009b) and have the advantage that inference based on these methods are valid, whether or not identification difficulties arise. In certain situations, particularly when some parameters are well identified and others are not, these methods allow for important insights.

Instrument selection

When setting up the moment conditions for the GMM approach, one has to be specific about the choice of instrument variables $\mathbf{z}$. Rational expectations models such as our Phillips curve specification suggest that all information available at the time when the forecast is made can be used as valid instruments. Because of simultaneity and potential publication lags, we consider only instruments dated time $t-1$ or earlier.

Since the potential instrument set contains the full information set of private agents, the number of candidate instruments is infinite. This really means that any measured past variable can serve as a potential instrument. Additionally, there is no theoretical justification or practical guide for deciding which instruments to use for estimation. However, early Monte Carlo experiments for GMM suggest that using as many instruments as possible is not a good idea. Instead, one should be somewhat selective when choosing $\mathbf{z}$ (see Tauchen, 1986; Kocherlakota, 1990). In empirical macroeconomic applications it has become standard practice for instruments to be chosen relatively unsystematically from among recent macroeconomic variables that are thought to predict the instrumented variables well.

The success of IV methods clearly depends on the quality of instruments, so it is logical to put some effort into careful selection of the instrument set in order to get more reliable results. We therefore adopt the Factor-GMM approach by Bai and Ng (2010) and Kapetanios and Marcellino (2010), who employ principle components obtained from a large data set as potentially relevant instruments. Based on the idea of Stock and Watson (2002) that the information of large datasets can be summarized by a few factors, those factors should reflect most of the private agents' information set. Additionally, Stock and Watson (1999) show that those factors are useful in forecasting inflation, which suggests that factors may be relevant instruments (since they are correlated with future inflation).

More precisely, let us assume that there are N potential instrumental variables x_t available and that these are generated by the factor model:

$$x_t = \Lambda' g_t + \nu_t, \tag{11}$$

where it is assumed that the number of static factors r is much smaller than N.[6] Consequently, the factors g_t are natural instrument choices. In addition to the

[6] As a decision rule, we consider all factors that explain at least 10% of the overall variance.

factors, we also allow key macroeconomic indicators and lags of the endogenous variables as part of the instrument set. As potential additional indicators, we consider variables used in other studies (see Galí and Gertler, 1999; Ravenna and Walsh, 2006).

The second step is to apply a general-to-specific modeling strategy to eliminate redundant instruments (those which correlate only marginally with the variables they instrument). This is carried out separately for each instrumented variable.[7] Those variables that remain in one of the equations enter into the final instrument set.[8] If only the most relevant instrumental variables are considered in the estimation process, it promotes identification robust methods on account of power gains. The exclusion of redundant instruments can thus help to avoid the well-documented power loss of the AR statistic when the number of instruments increases (Andrews and Stock, 2005).

Identification robust inference

Conducting inference of the parameters of interest relying on standard Wald-type tests and t-statistics is usually done in the standard GMM framework, which is problematic when identification difficulties occur. Typically, the GMM estimate is treated as if

$$\hat{\vartheta} \approx N(\vartheta_0, \hat{V}/T)$$

[7] The general-to-specific approach always deletes the least significant variable from the equation until all remaining p values are below a pre-specified threshold (which we set at 0.1).

[8] We also consider a system-based approach whereby all equations are considered jointly and a blockwise elimination is performed (until a certain information criterion is minimized). However, this approach results in very few instruments (sometimes with fewer instruments as instrumented variables), whereas the equation strategy leaves us with slightly more instrumental variables (7 indicators) which are more convenient to work with.

where

$$\hat{V} = \left(\hat{D}\ \hat{S}_T^{-1}\hat{D}'\right)^{-1} \quad \text{and} \quad \hat{D} = 1/T \sum_{t=1}^{T} \frac{\partial f_t(\vartheta)}{\partial \vartheta'}|_{\vartheta=\hat{\vartheta}} \tag{12}$$

and S_T is the long-run covariance matrix (see Hamilton, 1994, Chapter 14). From that, standard $(1-\alpha)100$ confidence intervals for the individual parameter i can be computed as

$$\hat{\vartheta}_i \pm z_{\alpha/2}\sqrt{\hat{V}_{ii}/T},$$

where $\hat{V}_{ii}$ is the $i - i^{th}$ element of the matrix $\hat{V}$ given in eq. (12) and $z_{\alpha/2}$ is the 100(1-$\alpha/2$) percentile of the standard normal distribution. One strategy increasingly applied in IV and GMM estimation is to pre-check whether the instruments are strong enough (see e.g., Hahn and Hausman, 2002; Stock and Yogo, 2005). If this is the case, standard methods of inference are applied as outlined above.

In contrast, this study adopts a different perspective and applies identification robust methods of inference. We concur with the arguments by Dufour (2003), Andrews and Stock (2005) and Kleibergen (2007), who have shown that using conventional inference methods after pretesting for identification is both unreliable and unnecessary. The unreliability resides in the fact that the size of such a two-step procedure cannot be controlled. The sequential procedure is unnecessary because identification robust methods are as powerful as the standard methods when instruments are strong and more powerful than the two-step procedures when instruments are weak (see Kleibergen and Mavroeidis, 2009b). Moreover, identification robust methods can be helpful only when certain parameters are unidentified or close to unidentified, while other parameters are well identified.

As shown by Mavroeidis (2005), identification of the NKPC for economic plausible parameter values is challenging, and weak instrument problems are very likely to occur. This view is supported empirically by numerous studies (Ma, 2002; Dufour et al., 2006; Mavroeidis, 2007; Martins and Gabriel, 2009; Kleibergen and Mavroeidis, 2009b; Nason and Smith, 2008) comparing weak instrument robust tests with standard Wald-type tests obtained with GMM. We essentially adopt their idea, and do not assume a priori that the parameters are identified. We used **S**-sets, as proposed by Stock and Wright (2000) and applied to the NKPC by Ma (2002) and Mavroeidis (2007), which can be constructed from the CUE objective function (see eq. 10). This method has similar important characteristics, described by Dufour (1997) as identification robust. This requires unbounded confidence intervals (thus uninformative) whenever parameters are unidentified. In the situation where parameters are weakly identified, this should translate into fairly large confidence sets. As shown by Dufour (1997), this is not the case with standard Wald-type methods, which hold only when identification is fully guaranteed and when there are no weak instrument problems. Other than this, these methods are unreliable and standard normal approximations provide a very unsatisfactory guide for inference.

The **S**-sets used for constructing confidence sets are very close to the well-known overidentification test of Anderson and Rubin (1949). Several authors (including e.g., Dufour, 1997; Stock et al., 2002; Dufour, 2003; Andrews and Stock, 2005; Dufour and Taamouti, 2005; Dufour and Taamouti, 2007) provide evidence that this statistic is fully robust to weak instrument problems. When it comes to linear simultaneous equation models, Stock and Wright (2000) have shown that **S**-sets are asymptotically equivalent to confidence sets obtained by inverting the Anderson-

Rubin (AR) statistic. So **S**-sets can be seen as an extension for the AR test in linear models to GMM as a more general model class. To obtain **S**-sets, that is, a joint confidence set for the parameter vector ϑ, we use Stock and Wright's (2000) result that $S(\vartheta_0) \stackrel{D}{\to} \chi^2_k$, where $S(\vartheta_0)$ is the CUE objective function eq. (10) evaluated at the true parameter values ϑ_0 and k is the number of instruments. The joint confidence interval consists of those parameter values for which the test statistic does not reject. Thus an asymptotically valid $100(1-\alpha)\%$ confidence interval for the parameter vector ϑ is given by

$$\{\theta : T \times \mathbf{S}(\vartheta) < c_k(\alpha)\}, \tag{13}$$

where $c_k(\alpha)$ denotes the $100(1-\alpha)\%$ percentile of the χ^2_k.[9]

This procedure can be applied to both the reduced form parameters γ_f, γ_b, λ and ϕ^i and the structural parameters β, θ, ξ and ϕ^i (given the calibrated values for α and ϵ). The resulting **S**-set is four-dimensional for the full model specification.[10]

Confidence intervals for the individual parameters are obtained by using the projection method. The idea is that projection-based tests do not reject the individual hypotheses $H_0 : \beta = \beta_0$ when the joint hypothesis $H^* : \beta = \beta_0, \alpha = \alpha_0$ do

[9]The construction of joint confidence intervals involves searching for values within an economically plausible range and collecting those values which the test does not reject. This is done by means of a grid search procedure.

[10]Additional methods are now available for dealing with weak instrument problems within the GMM setting (see Kleibergen and Mavroeidis, 2009b, for a comparison of different IV robust methods with application to the Phillips curve). Kleibergen and Mavroeidis (2009b) consider not only **S**-sets, but also a score Lagrange Multiplier (KLM) test, the difference between **S**-sets and the KLM statistics (JKLM) and an extension of the conditional likelihood ration test of Moreira (2003) to GMM (MQLR). Their simulation results indicate that the MQLR is at least as powerful as any of the other tests. However, while MQLR dominates the **S** statistics under some conditions in terms of power, it also imposes additional restrictions on the reduced form models and may consequently be more fragile. This could translate into problems when relevant instruments are missing (this point was raised by Dufour, 2009).

not reject for some values of α. This test method is proposed by Dufour (1997), Dufour and Jasiak (2001), Dufour and Taamouti (2005) and Dufour and Taamouti (2007). This procedure is fully robust to weak instruments but its drawback is that projection-based tests are conservative.[11]

A further characteristic of identification robust confidence intervals based on the CUE objective function is that they may be empty. This is the case when the test rejects for all possible parameter values. Thus, **S**-sets already include a test of overidentified restrictions comparable to a J test as proposed by Hansen. If no parameter vector is compatible with the specified model, the corresponding confidence sets will be empty. We interpret these results as a rejection of the empirical model.

An MCMC approach for calculating parameter uncertainty

We also present an alternative to the projection method to approximate the estimation uncertainty of the individual parameters. We therefore make use of simulation techniques as originally proposed in the Bayesian literature. This enables us to systematically characterize the shape of the GMM objective function in situations where non-linearities and dimensionality complicate traditional methods of inference (and where the projection technique is less powerful). We concur with Chernozhukov and Hong (2003) by constructing

[11] Another available approach can be applied to parameter subsets (see Stock and Wright, 2000; Kleibergen and Mavroeidis, 2009a). In this case, some parameters are assumed to be identified. As long as the assumptions are satisfied, these tests are asymptotically non-conservative and more powerful than projection-based tests. However, when estimating the structural parameters, the parameter space is bounded. This implies that the restricted estimates may fall on the boundary, which violates the conditions for subset tests (Kleibergen and Mavroeidis, 2009b).

$$\varphi_T(\vartheta) \propto \exp\left[-\frac{1}{2}\mathbf{S}(\vartheta)\right],$$

with $\mathbf{S}(\vartheta)$ the CUE objective function. The right-hand side function is scaled so that

$$\int \varphi_T(\vartheta)\mathrm{d}b = 1.$$

Now we make use of MCMC (Markov chain Monte Carlo) methods to summarize $\varphi_T(\vartheta)$ and hence $\mathbf{S}(\vartheta)$. Typically, MCMC methods are used in Bayesian analysis in conjunction with likelihood functions. In this application we use the CUE criterion function instead, and run random parameter searches to evaluate its properties (see Hansen, Heaton, Lee and Roussanov, 2007, for a similar application based on CUE on the intertemporal elasticity of substitution).[12]

Although we do not give a Bayesian interpretation, we may infer "marginals" for individual components of the parameter vector by averaging out the remaining components. The integration method contrasts with the standard practice of concentration inferior to standard methods of inference, where minimization and computing derivatives at minimized values are required. In applying the simulation technique based on the $\mathbf{S}$-sets we hope to gain more insight into the individual parameter uncertainty (since we get the whole distribution) than if we were to rely only on the projection-based confidence intervals.

[12]See Appendix A.2 for detailed information.

3.3 Selecting among candidate models

So far, we have concentrated on methods of inference for one particular model. However, in comparing different models in relation to one economic phenomenon, it is necessary to choose from among different candidate models. When one model is nested within another, it is possible to test for these parameter restrictions given the methods described above. This is, e.g. the case for deciding whether the Phillips curve model (eq. 9) is purely forward-looking. This translates into a test of $H_0: \ \xi = 0$.

However, if this is not the case, i.e. meaning that one model is not a special case of the other, the models are *non-nested* and additional methods have to be employed. Non-nested tests for GMM have been proposed by Singleton (1985) Ghysels and Hall (1990) and Smith (1992). Since none of these approaches is really satisfactory (Hall, 2005), we apply the relatively simple model selection criteria proposed by Andrews and Lu (2001). These information criteria can be seen as GMM counterparts of the likelihood information criteria (BIC, AIC,...) and are based on the J-test statistic for testing over-identifying restrictions (see Hansen, 1982).

Defining the moment and model selection criteria (MMSC) as

$$\text{MMSC}_T = J_n(b) - h(k - |b|)\kappa_T, \tag{14}$$

where $J_T(b)$ is the test statistic of the over-identification test, given parameters b. Let $|b|$ denote the number of parameters to be estimated given b and k the number of moment conditions. Comparable to likelihood based criteria $h(k - |b|)\kappa_T$ is a "bonus term" that penalizes the increasing number of estimated parameters and rewards the utilization of more over-identifying restrictions.

Standard examples of MMSC are the BIC and AIC criteria for model selection. Those are defined as

$$\text{MMSC-BIC:} \quad \kappa_T = \ln T \text{ and } \text{MMSC}_{BIC,T} = J_n(b) - (k - |b|) \ln T,$$

$$\text{MMSC-AIC:} \quad \kappa_T = 2 \text{ and } \text{MMSC}_{AIC,T} = J_n(b) - 2(k - |b|).$$

Typically, the MMSC measures describe a trade-off between the magnitude of the J statistic and the number of parameters (and moment conditions) employed. The model with the lowest MMSC value is to be preferred. As shown by Andrews and Lu (2001), the selection procedures can help to specify a model, where the MMSC-BIC in particular is found to work quite well for this purpose.

3.4 Data

We use quarterly time series data with a sample period ranging from 1960q1-2005q4 to estimate eq. (9) for the U.S. economy. The data are taken from the OECD Quarterly National Accounts database, IMFs International Financial Statistics (IFS) and the indicator database provided on Mark Watson's homepage. Inflation is defined as the quarterly log difference of the GDP deflator. Real marginal cost is proxied by the labor share of output, which is defined as the ratio of total compensation to nominal GDP. As a measure for the short-run nominal interest rate, two definitions are considered: 3-month Treasury bill rates and bank lending rates.[13] Both explanatory variables – labor share and interest rates – are defined as percentage deviations of a steady state value, while inflation rate is expressed as percentage point deviations.

The potential instrument set is composed of lags of inflation, the labor share and short-term interest rates (up to four lags are considered). Additional instruments consist of a yield spread, $(r^l - r^m)_t$, defined as the 10-year government bond yield minus the 3-month Treasury bill rate, wage inflation Δw_t and a quasi-real time detrended output gap $\tilde{y}_t$ (which is computed recursively and contains information only up to period t). These additional instruments contain up to two lags in the instrument selection step. The factors are extracted using the Stock and Watson database covering 108 macroeconomic variables over the full sample period. We take the four principle components $(g^1, ..., g^4)$, which provide the largest explanatory power (at least 10% of the overall variance) as candidate instruments. The step-wise selection approach applied for each instrumented variable results in 17 selected instrument variables (from a candidate set of 24 candidates) plus a constant. This instrument set includes

$$\begin{aligned}\mathbf{z}^{opt}_{t-1} = [c\ \hat{\pi}_{t-1}\ \hat{\pi}_{t-2}\ \hat{\pi}_{t-3}\ \hat{\pi}_{t-4}\ \hat{s}_{t-1}\ \hat{R}^i_{t-1}\ \hat{R}^i_{t-2}\ \hat{R}^i_{t-3}\ \hat{R}^i_{t-4} \\ (r^l - r^m)_{t-1}\ \tilde{y}_{t-2}\ g^1_{t-1}\ g^1_{t-2}\ g^2_{t-1}\ g^2_{t-2}\ g^3_{t-1}\ g^3_{t-2}]'.\end{aligned} \tag{15}$$

As a robustness check we also consider an instrument set similar to Ravenna and Walsh's (2006) set, including

[13]Bank lending rates for the U.S. are taken from the IMFs International Financial Statistics (IFS).

$$\mathbf{z}_{t-1}^{rw} = [c\ \hat{\pi}_{t-1}\ \hat{\pi}_{t-2}\ \hat{\pi}_{t-3}\ \hat{\pi}_{t-4}\ \hat{s}_{t-1}\ \hat{s}_{t-2}\ \hat{R}_{t-1}^{i}\ \hat{R}_{t-2}^{i}\ \hat{R}_{t-3}^{i}\ \hat{R}_{t-4}^{i} \\ \Delta w_{t-1}\ \Delta w_{t-2}\ \tilde{y}_{t-1}\ \tilde{y}_{t-2}\ (r^{l}-r^{m})_{t-1}\ (r^{l}-r^{m})_{t-2}]'. \tag{16}$$

The main differences between these two sets is the use of factors in the instrument set $\mathbf{z}_{t-1}^{rw}$ from eq. (15). The size is similar.

4 Estimation results

In the following, we present the baseline results for the econometric model as outlined in section 3. We therefore start by considering the most general specification (denoted as I_{Full}), including the cost channel measured by the Treasury bill rate (in line with Chowdhury et al., 2006; Ravenna and Walsh, 2006). We assume partial indexation (measured by ξ) as well as the existence of real rigidities (reflected in the calibrated term $\frac{1-\alpha}{1+\alpha(\epsilon-1)}$). Results are displayed only for the structural form, where structural parameters are estimated directly from the non-linear equation. The advantage of this is that economically relevant quantities become directly apparent. In particular, the degree of nominal rigidities, reflected in θ and the corresponding average frequency of price reoptimization (or adjustment), can be easily deduced. However, the reduced form coefficients can be easily computed once the structural parameters have been determined. So one can compare the results directly to Chowdhury et al.'s (2006) findings. They used the same explanatory variables but estimated the linear form.[14]

[14] Owing to the application of the CUE (instead of the 2GMM) the point estimates of the directly estimated reduced form coefficients would be exactly the same as those recalculated from the structural parameters.

Table 1 provides the CUE point estimates together with identification robust 90% confidence intervals (**S**-Sets) calculated by using the projection method. For comparison, we also display standard Wald-type confidence intervals with the same level of significance. Further, the average frequency of price reoptimization is calculated from the Calvo parameter θ together with its confidence interval. In addition, p-values associated with the CUE point estimates are provided, which can be interpreted in the same way as the standard test for overidentifying restrictions.[15] To compare different model specifications (which may or may not be nested), we present the GMM information criterion MMSC-BIC for each model. As outlined above, models with lower MMSC-BIC values are to be preferred. All specifications presented in Table 1 are based on the same degree of real rigidity; namely $\alpha = 0.27$ (the firm specific capital share) and $\epsilon = 11$ (which implies a steady state mark-up of 10%). Further, the West (1997) HAC estimator with an MA(1) process is used in estimation.

Turning to the results for the full specification I_{Full}, we obtain a CUE parameter vector for ($\hat{\beta}$, $\hat{\theta}$, $\hat{\xi}$, $\hat{\phi}^m$) of $(0.97, 0.67, 0.16, 1.39)$. The estimates for β, θ and ϕ^m are very close to Ravenna and Walsh's (2006) results, although we allow for partial indexation (which provides a rationale for including a lagged inflation term) and we used a different instrument set. The implied average frequency of price adjustment is roughly three quarters, which is basically in line with the literature.

[15]Note that the CUE point estimates can also be interpreted as Hodges-Lehmann estimators (which are the least-rejected values and can be interpreted as point estimates). It may be the case that the CUE objective function evaluated at the least-rejected values (the CUE estimates) exceeds the $100(1-\alpha)$ percentile of the χ^2_k. This case results in an empty confidence set (this happens when the associated p-value is below 0.1) and can be interpreted as a rejection of the model rather like the Hansen test of overidentification. The main difference between the two is that the Hansen test employs a χ^2_{k-s} distribution, with $k-s$ degrees of freedom (where s is the number of estimated parameters), as opposed to k.

Table 1: Estimates of the structural parameters

Model specification:

$$\hat{\pi}_t = \frac{\beta}{1+\beta\xi}\hat{\pi}_{t+1} + \frac{\xi}{1+\beta\xi}\hat{\pi}_{t-1}\frac{(1-\theta\beta)(1-\theta)}{(1+\beta\xi)\theta}\frac{1-\alpha}{1+\alpha(\epsilon-1)}\left(\hat{s}_t + \phi_i \hat{R}_t^i\right) + u_t$$

	$\mathbf{I}_{\text{Full}}$	$\mathbf{II}_{\text{NKPC}}$	$\mathbf{III}_{\text{RW}}$	$\mathbf{IV}_{LR}$	$\mathbf{V}_{\text{RW-LR}}$
β	0.9706	1.0110	0.9793	0.9637	0.9645
	(0.92,1.03)	(0.96,1.06)	(0.93,1.03)	(0.92,1.01)	(0.92,1.01)
	[0.84,1.00]	[0.92,1.00]	[0.84,1.00]	[0.88,1.00]	[0.86,1.00]
θ	0.6679	0.6212	0.6369	0.6327	0.6184
	(0.47,0.87)	(0.44,0.80)	(0.53,0.74)	(0.58,0.68)	(0.56 ,0.68)
	[0.45,1.00]	[0.45,1.00]	[0.42,1.00]	[0.55,0.95]	[0.55,1.00]
ξ	0.1621	0.1138	0	0.2067	0
	(0.01,0.31)	(-0.08,0.21)		(0.05,0.36)	
	[0.00, 0.60]	[0.00,0.70]		[0.00,0.65]	
ϕ^m	1.3888	–	0.8828	–	–
	(-1.04,3.82)		(0.86,0.90)		
	$[-\infty,\infty]$		$[-\infty,\infty]$		
ϕ^l	–	0		1	1
Implied Freq.	3.01	2.64	2.75	2.72	2.62
$1/(1-\theta)$	$[1.82,\infty]$	$[1.82,\infty]$	$[1.82,\infty]$	[2.13,20]	$[2.00,\infty]$
p-Value	0.4423	0.2840	0.4013	0.4534	0.3799
MMSC-BIC	-58.58	-56.90	-58.22	-60.38	**-63.80**

Notes: Point estimates are obtained using CUE. 90% projection based confidence intervals in squared brackets and 90% Wald confidence intervals in round brackets. P-values report the test for the joint confidence set evaluated at the CUE point estimates. West's (1997) HAC estimate is used. Sample period: 1960:1-2005:4. Factor-augmented and preselected instrument set (see eq. 15) is used.

However, when turning to the uncertainty of the estimates reflected by the **S**-sets, it is obvious that the length of the confidence interval for ϕ^m is infinite.[16] This extreme finding results from the fact that we cannot rule out the case of $\theta = 1$. Economically, this is the case of total price rigidity where prices are never reoptimized (which also results in a Phillips curve slope of zero). Provided that $\theta = 1$, it also follows that $\frac{(1-\theta\beta)(1-\theta)}{(1+\beta\xi)\theta}\frac{1-\alpha}{1+\alpha(\epsilon-1)} = 0$, which implies that ϕ^m may take

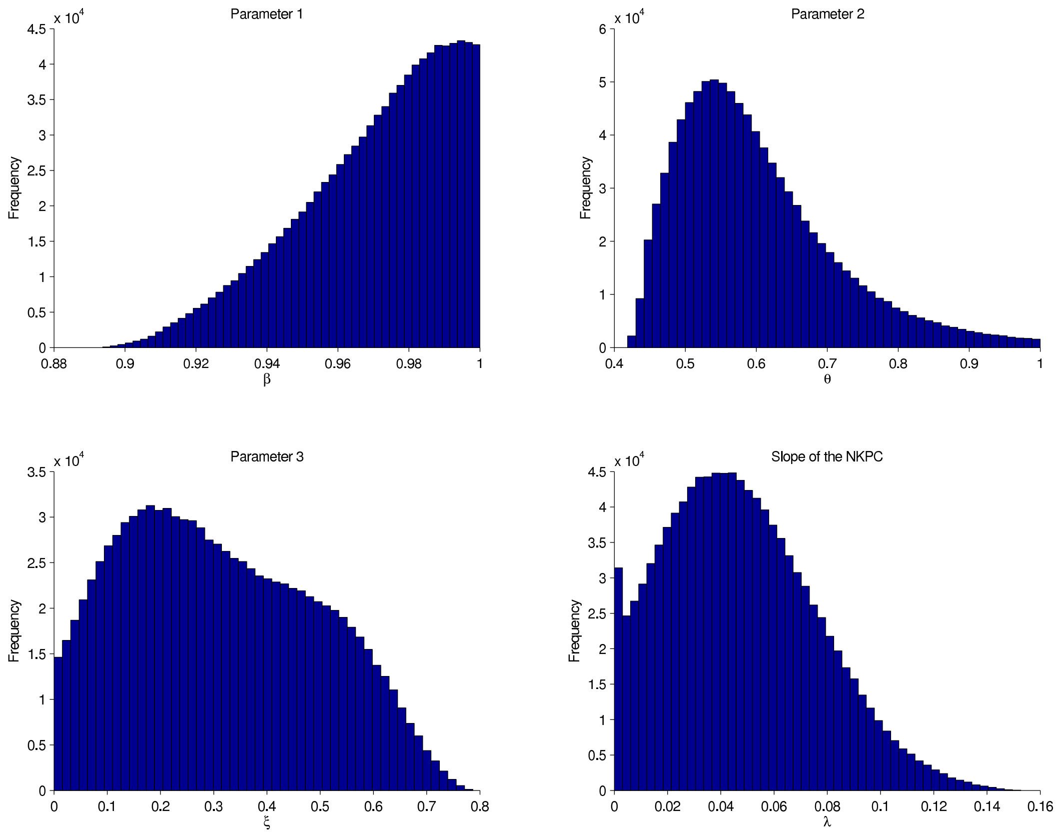
[16] Note that a confidence interval of infinity reflects the fact that this parameter is unidentified (given the model structure and the data). As emphasized by Dufour (1997), this feature of allowing for unbounded confidence intervals whenever parameters are unidentified is exactly what characterizes identification robust procedures.

any value. This obviously leads to an unidentifiable parameter ϕ^m. Identification of ϕ^m thus requires $\theta = 1$ not to be part of the confidence set, so that $H_0: \ \theta = 1$ can be rejected. Obviously this is not guaranteed in specification I_{Full}. It also implies that ϕ^m is unidentified in the linearized version of the model, where, if $\lambda = 0$ cannot be rejected, ϕ^m cannot be identified. Thus, it can be also seen that the model structure estimated by Chowdhury et al. (2006) is confronted by the same identification difficulties as the structural model, at least for the U.S. economy.

Typically, Wald-type confidence intervals turn out to be smaller in comparison with identification robust sets and may not detect the identification problem of ϕ_m, even when the Wald confidence interval for θ would include the case $\theta = 1$. Other studies using identification robust methods for evaluating the standard NKPC (without considering a cost channel) likewise cannot rule out the case of total nominal price rigidity ($\theta = 1$) which also translates into an identification difficulty of ϕ^m (see e.g., Ma, 2002; Kleibergen and Mavroeidis, 2009b). Despite the identification problems for ϕ_m, other parameters can still be analyzed. For the measure of partial indexation ξ, we can conclude that it is basically unimportant and not significantly different from zero. In contrast, the case of full indexation, which implies $\xi = 1$ and has been assumed by Christiano et al. (2005) and Eichenbaum and Fisher (2007) can be clearly rejected. The confidence interval for β is relatively tight and is in an economically plausible range of something close to 1.

Although we demonstrated that problems exist for parameter ϕ^m identification and that the model structure is more or less useless as far as revealing any conclusion about the existence of a cost channel goes, we proceed with alternative specifications in order to assess the importance of the cost channel effect. We therefore take a standard Phillips curve specification with the restriction $\phi^i = 0$ as a benchmark to see whether different restrictions (with or without the cost channel) are to be preferred relative to a standard NKPC. The results for the standard Phillips curve are given by specification II_{NKPC} and the findings are in line with those of Kleibergen and Mavroeidis (2009b). Similar to specification I_{Full}, we get a rather large confidence interval for ξ, where we cannot reject the case of $\xi = 0$. This means that either indexation seems to play a minor role in inflation dynamics or that the lagged inflation term is relatively unimportant. However, owing to the large confidence intervals there is substantial uncertainty. We can again reject the null of full indexation. This is in contrast to Kleibergen and Mavroeidis's (2009b) findings, which point to an uninformative confidence interval for the backward-looking component which consists of the entire parameter space. The difference may be attributed to the slightly different model specification (we assume partial indexation, whereas Kleibergen and Mavroeidis assumes rule-of-thumb firms). Even more important may be the use of factors as instruments, which seems to make identification easier. This is in line with Wright's (2009) argument that the choice of instruments should be conducted according to principles of forecasting. He illustrated this by using better predictors for inflation and the labor share (namely inflation expectations from surveys) as instruments and obtaining smaller confidence sets, as do Kleibergen and Mavroeidis (2009b). Since factors have been successfully used to forecast output and inflation, the same principle may apply here. However, for the Calvo parameter θ, we cannot reject total nominal price rigidities because the confidence interval for θ includes 1 and the confidence interval for the average frequency of price adjustment is unbounded from above.

Figure 1: MCMC with continuously-updated GMM criterion function: with $\phi_l = 0$



Notes: The histograms contain 1 million valid draws of the MCMC algorithm as described in appendix A.2. The parameters β, θ and ξ are restricted to the interval $[0, 1]$ and to be inside the joint **S**-set.

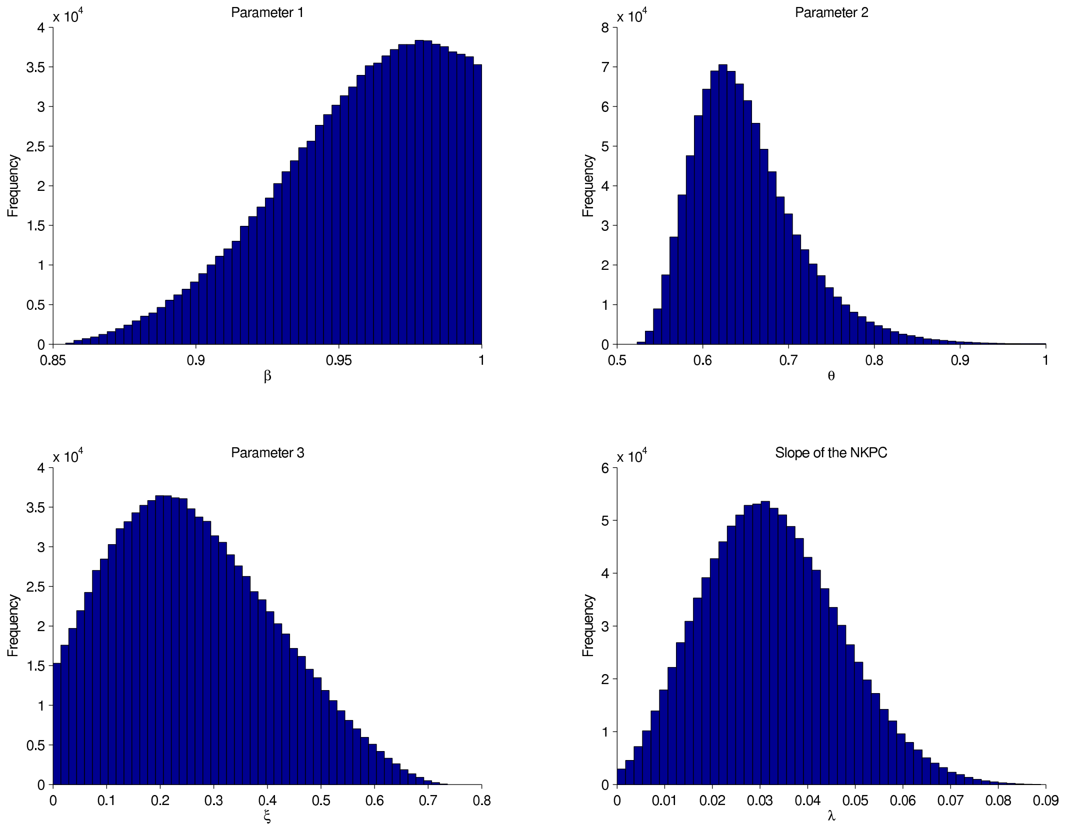
Turning to specification III_{RW}, this is essentially the estimated model of Ravenna and Walsh (2006), which omits the lagged inflation term and assumes static indexation ($\xi = 0$). Given this model, the same identification difficulties for ϕ^m arise as in specification I_{Full}. The restriction of $\xi = 0$ does not help in identifying parameter ϕ^m, and standard test procedure would lead to an entirely incorrect conclusion, namely a significant cost channel parameter ϕ^m, associated with low estimation uncertainty. The results for the remaining coefficients are similar to I_{Full} and II_{NKPC}. However, it is remarkable that the MMSC-BIC criterion favors III_{NKPC} over II_{NKPC}, providing additional evidence that partial indexation is unimportant in our setting.

If we employ a more direct measure of short-term liabilities, namely the lending rate (as proposed by Tillmann, 2008) instead of the Treasury bill rate, we can make use of the restriction that $\phi^l = 1$. This implies that changes in the lending rate fully translate into firms' marginal costs. This follows naturally from the eqs. (3, 7 and 8). Specifications IV_{LR} and V_{LR-RW} make use of this fact. Although the point estimates remain roughly the same in comparison with previous specifications, it turns out that the latter are to be preferred, according to the model selection criterion. Thus a model including the cost channel is clearly preferable. Moreover, we can now reject the case of full price rigidity $\theta = 1$, which can be also interpreted in favor of the model including the cost channel. This further implies that the confidence interval for the average frequency of price optimization is now bounded from above. In addition, it supports the omission of the lagged inflation term given, that the hypothesis $\xi = 0$ cannot be rejected. This is remarkable, because many studies have criticized pure forward-looking Phillips curve specifications on account of their inability to provide enough inflation persistence.

To strengthen our results, we make use of the MCMC approach (see section 3.2 and appendix A.2) to get a more complete picture of parameter uncertainty, given different Phillips curve specifications, and to find out whether or not those provide evidence in favor of a cost channel effect. As a first attempt, we simulate the parameter uncertainty associated with specification II_{NKPC} and IV_{LR}, which implies different calibrations for the parameter ϕ^l.[17] The simulation reveals differ-

[17] In principle, one can also simulate the full model specification I_{Full}. However, owing to the severe identification difficulties of parameter ϕ^m, this is extremely cumbersome. Additionally, the results depend heavily on the tolerated parameter space for ϕ^m. However, when simulating

Figure 2: MCMC with continuously-updated GMM criterion function: with $\phi_l = 1$



Notes: The histograms contain 1 million valid draws of the MCMC algorithm as described in appendix A.2. The parameters β, θ and ξ are restricted to the interval $[0, 1]$ and to be inside the joint **S**-set.

ent shapes for the structural parameters β, θ and ξ and the consequences for the Phillips curve slope λ (see Figure 1 and 2). Overall, the parameter uncertainty seems to be smaller in the case of $\phi^l = 1$. Particularly, the distribution of θ is much more centered, involving less uncertainty, and it is less skewed in case of $\phi^l = 1$. Even more important is the fact that there is no probability mass close to $\theta = 1$, which implies that the extreme case of perfect nominal price rigidity is quite unlikely. Instead, most of the probability mass lies between 0.55 and

from the full model an ill-behaved distribution is obtained for parameter ϕ^m characterized by multimodality and a large overall dispersion.

0.85. Further, the simulated density for ξ looks more standard with the restriction $\phi^l = 1$. In both cases, full indexation ($\xi = 1$) can be rejected. For the standard NKPC ($\phi^l = 0$), the density is strongly skewed, still with a substantial probability mass for $\xi > 0.5$. Only by considering the cost channel $\phi^l = 1$ can the uncertainty about the slope of the Phillips curves be characterized by a symmetric distribution located with high confidence in the positive parameter space. In contrast, with $\phi^l = 0$, the dispersion for λ is larger and the distribution is pulled towards zero, which results from the greater uncertainty of θ in the case of the NKPC. Generally, we interpret all these findings as evidence in favor of a cost channel.

Robustness

Having set up the baseline model estimates for evaluating the cost channel along the lines of a Phillips curve model, we now check the robustness of the results for specific modifications of the standard model. We tackle the issues of further parameter restrictions, an alternative HAC estimator, using different instrument sets (one of them similar to the set of Ravenna and Walsh and another which excludes the information in period $t-1$), a pure forward-looking Phillips curve specification excluding the cost channel (see Table 2), and sensitivity with respect to the different calibrations that determine the degree of real rigidity.

First, we naturally employ the restriction $\beta = 0.99$ and check whether this can help in identifying the remaining parameters (see Table 2 specification VI and VII). While the additional restriction does not provide much towards estimating ϕ^m, which is still unidentified (see VI), it does give some assistance in identifying θ in specification VII. Together with the restriction on the discount rate, we can

Table 2: Sensitivity Analysis

Model specification:

$$\hat{\pi}_t = \frac{\beta}{1+\beta\xi}\hat{\pi}_{t+1} + \frac{\xi}{1+\beta\xi}\hat{\pi}_{t-1}\frac{(1-\theta\beta)(1-\theta)}{(1+\beta\xi)\theta}\frac{1-\alpha}{1+\alpha(\epsilon-1)}\left(\hat{s}_t + \phi_i\hat{R}_t^i\right) + u_t$$

	VI	**VII**	**VIII**	**IX**	**X**	**XI**
β	**0.99**	**0.99**	0.9594	0.9485	0.7539	1.0063
			(0.90,1.02)	(0.91,0.99)	(0.52,0.98)	(0.95,1.06)
			[0.82,1.00]	[0.78,1.00]	[0.00,1.00]	[0.92,1.00]
θ	0.6533	0.6490	0.6445	0.6605	0.5930	0.6115
	(0.59,0.71)	(0.60, 0.69)	(0.56,0.73)	(0.62,0.70)	(0.47,0.72)	(0.43,0.79)
	[0.50,1.00]	[0.60,0.85]	[0.50,1.00]	[0.60,0.90]	[0.25,1.00]	[0.45,1.00]
ξ	0	0.1989	0.2659	0.3712	0.5361	**0**
		(0.14,0.25)	(0.10,0.43)	(0.23,0.51)	(0.32,0.74)	
		[0.00,0.65]	[0.00,0.65]	[0.00,0.90]	[0.15,0.95]	
ϕ^m	0.8406	–	–	–	–	–
	(0.71,0.96)					
	[$-\infty$,∞]					
ϕ^l	–	1	1	1	1	**0**
Implied Freq.	2.88	2.85	2.81	2.95	2.46	2.57
$1/(1-\theta)$	[2.00,∞]	[2.50,6.67]	[2,∞]	[2.5,10]	[1.33,∞]	[1.75,∞]
Instr.	opt	opt	opt	**rw**	**lag**	opt
HAC	West	West	**NW**	West	NW	West
p-Value	0.3280	0.4584	0.5451	0.6882	0.8177	0.2564
Wright-Test	$\infty^{\dagger}$	$5.02^{\dagger}$	1.25	$2.19^{\dagger}$	1.25	$1.90^{\dagger}$
MMSC-BIC	-62.91	-65.04	-61.13	-58.92	-17.09	-61.53

Notes: Point estimates are obtained using CUE. 90% projection based confidence intervals in squared brackets and 90% Wald confidence intervals in round brackets. P-values report the test for the joint confidence set evaluated at the CUE point estimates. West refers to West's (1997) HAC estimate and NW denote a Newey and West (1987) HAC estimator using 6 lags. Sample period: 1960:1-2005:4. Different instrument sets are used: opt (Factor-augmented and preselected instrument set) rw (the instrument set of Ravenna and Walsh) and lag (using only preselected variables and factors dated $t-2$ or ealier). The statistic of Wright is provided where † denotes significance at the 5% level.

reduce the uncertainty surrounding θ and the implied average frequency of price adjustment. The point estimates and the uncertainty associated with parameter ξ are in line with the unrestricted ones (see III_{RW} Table 2).

Second, we replace West's (1997) MA HAC estimator with the standard Newey-West estimator with a fixed bandwidth ($q = 6$) of the Bartlett kernel (specification

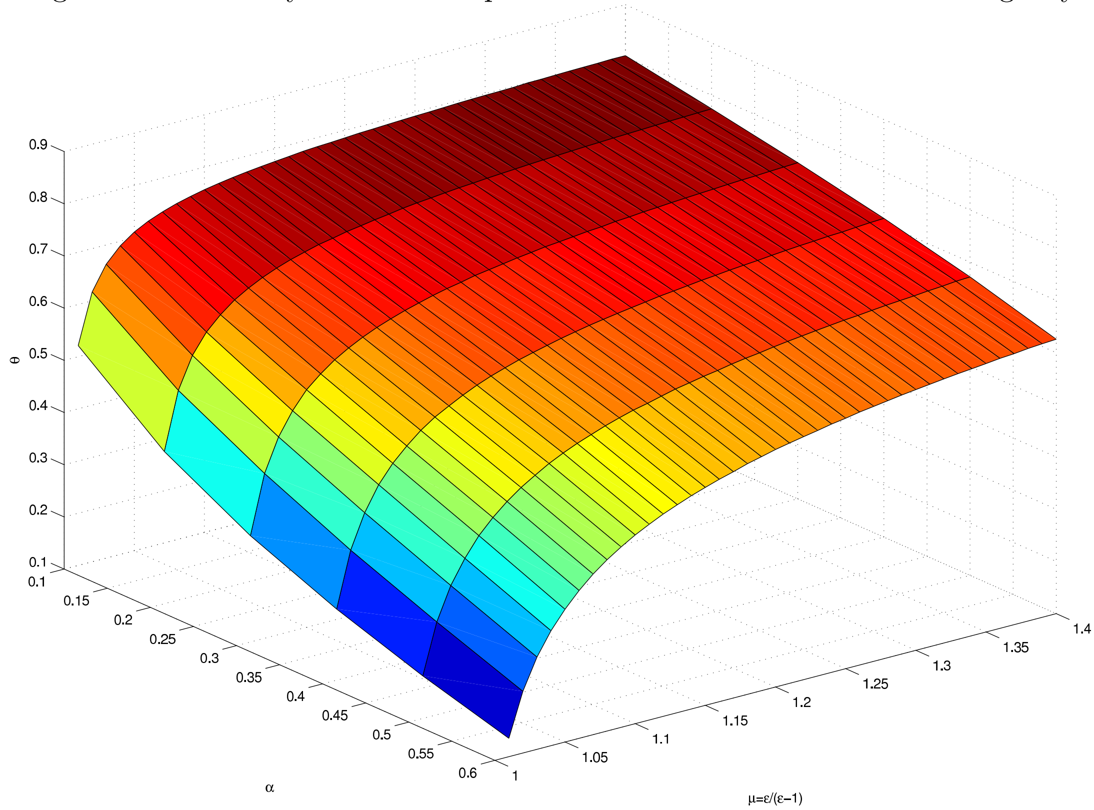
VIII). This hardly affects any of the previous conclusions. Switching between the two therefore does not affect the result. This may be also be interpreted as evidence that the residuals do not exhibit any substantial additional form of autocorrelation that would drive a wedge between the two estimators.

Third, we vary the instrument set. Ravenna and Walsh's (2006) instrument set B (labeled as $\mathbf{z}_{t-1}^{rw}$) is used as a comparison. This task results in a slightly smaller confidence interval for θ (larger values above 0.9 can now be rejected). However, at the same time, the confidence interval for ξ becomes much larger and contains nearly all the possible values – except for cases close to full indexation.[18] In addition, we try to take into account the possibility that the error term of the Phillips curve contains a cost-push shock that might be autocorrelated. Kuester, Müller and Stölting (2009) show that in this case the validity of moment conditions is violated. Namely, the error term is correlated with the instruments, and θ will be biased upwards. To check whether this could be the case in our setup, we use an instrument set containing only variables dated $t-2$ or earlier (so we exclude all instruments dated in $t-1$). If an autocorrelated cost-push shock is present, this will mitigate the bias effect (since correlation between the error term and the instruments should be smaller). When we observe a large deviation of the point estimates conditioned on the instrument set, we can interpret this as evidence in favor of an autocorrelated cost-push shock. (This would imply that all previous results are biased on account of this). From specification X, it follows that omitting the first lag in the instrument set does not change any of the results.

Fourth, when we choose a pure forward-looking Phillips curve specification

[18]Figure 5 in the appendix show bivariate plots for θ and ξ (given $\beta = 0.99$ and $\phi^l = 1$), depending on the two instrument sets.

Figure 3: Sensitivity of θ with respect to different calibrations of real rigidity



Notes: Different combinations of the calibrated values α and $\mu = \frac{\epsilon}{\epsilon - 1}$ are used for estimating θ, given the values $\xi = 0$ and $\beta = 0.99$.

without a cost channel given by XI (and the restriction $\xi = 0$ and $\phi^l = 0$) then the MMSC-BIC gives a higher value than specification V_{RW-LR}. Once again, this can be interpreted as evidence in favor of the cost channel. For all the additional results, we provide a test of adequacy of conventional asymptotics recently proposed by Wright (2010). The idea is thus simply to compare the volume of the robust confidence interval (**S**-set) with those of Wald confidence intervals obtained from the 2GMM estimator (see appendix A.3 for details). If this test is significant, standard asymptotics will be inappropriate (critical values are provided by the author). The results for the different specifications are displayed in Table 2.

It transpires that the test rejects in four out of six cases.[19] This points to the necessity for using identification robust methods.

Another important issue is the calibration of the degree of real rigidities in the U.S. economy. Since real rigidities in this model are empirically indistinguishable from nominal rigidities, we experimented with different calibrations to robustify our findings. From the production function and the assumption that the firms' specific capital stock is fixed in the short run, it follows that the degree of real rigidities depends on the capital share α and the constant demand elasticity ϵ. We investigate different calibrations and their effect on the estimated degree of nominal price rigidity (parameter θ). Since the steady-state markup rate is equal to $\mu = \frac{\epsilon}{\epsilon - 1}$, we consider economically plausible mark-up rates of between 1 and 40 percent. Although there is generally more consensus on the capital share (around 1/3), we compute calibrations of α between 0.1 and 0.6.[20] Figure 3 plots the estimated Calvo parameter θ, depending on different calibrations. It turns out that only with a high capital share and a high elasticity of demand ($\epsilon \to \infty$), which is equivalent to a very low mark-up, the estimated θ will be considerably smaller in comparison with the baseline estimates. On the other hand, with a smaller α and a larger mark-up, θ will be pulled upwards, but rarely above 0.8. Generally, we find that it needs large (and economically implausible) deviation from the baseline calibrations, particularly for the capital share, to affect the estimated Calvo-parameter to any great extent.

[19]Note that in this setting we have bounded the confidence set for the **S**-sets for the structural parameters in an economically plausible range between 0 and 1, while we leave the Wald confidence set unbounded. This may result in an underrejection of the test.

[20]e.g., Coenen, Levin and Christoffel (2007) note that α may be higher than normally expected to match the estimated degree of real rigidities.

Table 3: Error Measures

	IV	**XI**	**RW**	**AR(1)**
RMSE	**0.9596**	0.9729	1.2104	1.1714
MAE	**0.7576**	0.7655	0.9207	0.8827

Fit

For the sake of completeness, we also evaluate the fit of the Phillips curve specifications to see how the models track the data on actual inflation. We therefore concentrate on forward-looking specifications and compare the models employing different marginal cost definitions – the standard version with the labor share and the cost channel version (labor share plus lending rate).

Operationally, for comparing the model prediction with the actual inflation rate, we need a measure for the expected inflation rate. As in the 2SLS approach, we run a first-stage regression where realized future inflation rates are regressed on the instruments:

$$\hat{\pi}_{t+1} = \mathbf{z}_{t-1}^{'}\gamma + \nu_{t+1},$$

where the vector of coefficients γ is simply estimated by OLS. The predicted inflation rates are given by

$$\hat{\pi}_{t+1}^{p} = \mathbf{z}_{t-1}^{'}\hat{\gamma}$$

and $\hat{\pi}_{t+1}^{p}$ is the direct measure for inflation expectations which can be plugged in for $E_t\hat{\pi}_{t+1}$ in the structural model, given the estimated coefficients β and θ obtained from the GMM estimate.

It is now possible to compare the specifications employing different measures of marginal cost ($\hat{s}$ vs. $\hat{s}+\hat{R}^l$). Figure 4 shows the actual inflation rates against those predicted, assuming the existence of a cost channel (point estimates from specification V are used). Since graphically the two Phillips curve versions are difficult to distinguish from each other, we compute in-sample measures of fit, namely the root mean squared error (RMSE) and the mean absolute error (MAE) for both versions. In addition, we calculate the error terms for two simple benchmark models: a random walk (RW) model $\hat{\pi}_t = \hat{\pi}_{t-1} + v_t$ and an AR(1) model $\hat{\pi}_t = \delta\hat{\pi}_{t-1} + \eta_t$. As expected from the results based on MMSC-BIC, the model including the cost channel dominates the pure labor share-based Phillips curve version as well as the simple univariate time series models.

Conclusion

We find that the frequently applied test for the existence of the cost channel, i.e. $\phi^m = 0$, is inappropriate, because this parameter is basically unidentifiable, given the standard model structure. For this reason, there is not a great deal to be learned by relying on this practice, since robust confidence intervals should be uninformative in this case. This follows automatically with the application of identification robust econometric methods. This paper therefore compares different model specifications (which are non-nested) and assesses whether the marginal cost variable composed of the labor share and the bank lending rate (the model with cost channel) is more compatible with the data in comparison with the standard labor share version of the Phillips curve.

We use various techniques to gain a broader perspective on the empirical model.

First, we use factors as additional instruments to efficiently characterize the agent's information set, as implied by rational expectations in a parsimonious way. Then we apply identification robust techniques to evaluate the structural parameters of the model. Therefore, Stock and Wright's (2000) **S**-sets are used in combination with the projection method to get individual confidence sets that guarantee a high degree of robustness. We also apply an MCMC to characterize the uncertainty of the estimated parameters more efficiently. In addition, we use an information criterion (MMSC-BIC), which allows us to compare different model specifications (nested or non-nested) and which can be used to assess the relative model quality.

Our results reveal that a cost channel-based Phillips curve version is empirically more compatible with the data and with the theoretical arguments than the standard Phillips curve version. Moreover, in nearly all specifications, we cannot reject the fact that the degree of indexation (given by $\xi = 0$) is zero. This implies that the backward-looking element of inflation seems rather unimportant. Generally, a pure forward looking cost channel version of the Phillips curve characterizes inflation dynamics well. The direct effect of bank lending rates on firms' marginal costs may also account for the observed persistence of inflation rates beyond what can be explained by the standard NKPC.

Thus, it can be concluded that a pure forward-looking interest rate augmented Phillips is most compatible with the data. This suggests that considering the cost channel is an important aspect for monetary policy.

References

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

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation, *Econometrica* **59**(3): 817–859.

Andrews, D. W. K. and Lu, B. (2001). Consistent model and moment selection procedures for GMM estimation with application to dynamic panel data models, *Journal of Econometrics* **101**(1): 123–164.

Andrews, D. W. K. and Monahan, J. C. (1992). An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator, *Econometrica* **60**(4): 953–966.

Andrews, D. W. K. and Stock, J. H. (2005). Inference with weak instruments, *NBER Working Paper T0313*, NBER.

Bai, J. and Ng, S. (2010). IV-Estimation in a data-rich environment, *Econometric Theory* **26**(6): 1577–1606.

Brockwell, P. J. and Davis, R. A. (1990). *Time Series: Theory and Methods*, 2nd edn, Springer.

Calvo, G. A. (1983). Staggered prices in a utility-maximizing framework, *Journal of Monetary Economics* **12**(3): 383–398.

Canova, F. and Sala, L. (2009). Back to square one: Identification issues in DSGE models, *Journal of Monetary Economics* **56**(4): 431–449.

Chernozhukov, V. and Hong, H. (2003). An MCMC approach to classical estimation, *Journal of Econometrics* **115**(2): 293–346.

Chowdhury, I., Hoffmann, M. and Schabert, A. (2006). Inflation dynamics and the cost channel of monetary transmission, *European Economic Review* **50**(4): 995–1016.

Christiano, L. J., Eichenbaum, M. and Evans, C. (2005). Nominal rigidities and the dynamic effects of a shock to monetary policy, *Journal of Political Economy* **113**(1): 1–45.

Coenen, G., Levin, A. T. and Christoffel, K. (2007). Identifying the influences of nominal and real rigidities in aggregate price-setting behavior, *Journal of Monetary Economics* **54**(8): 2439–2466.

Den Haan, W. J. and Levin, A. (1997). A practitioner's guide to robust covariance matrix estimation, *in Handbook of Statistics: Robust Inference (Vol 15)*, North-Holland, pp. 291–341.

Dufour, J.-M. (1997). Some implausibility theorems in econometrics with applications to structural and dynamic Models, *Econometrica* **65**(6): 1365–1387.

Dufour, J.-M. (2003). Identification, weak instruments, and statistical inference in econometrics, *Canadian Journal of Economics* **36**(4): 767–808.

Dufour, J.-M. (2009). Comments on "Weak instrument robust tests in GMM and the New Keynesain Phillips curve" by F. Kleibergen and S. Mavroeidis, *Journal of Business & Economic Statistics* **27**(3): 318–321.

Dufour, J.-M. and Jasiak, J. (2001). Finite sample limited information inference in econometrics, *International Economic Review* **42**(3): 815–843.

Dufour, J.-M., Khalaf, L. and Kichian, M. (2006). Inflation dynamics and the New Keynesian Phillips curve: An identification robust econometric analysis, *Journal of Economic Dynamics & Control* **30**(9-10): 1707–1727.

Dufour, J.-M. and Taamouti, M. (2005). Projection-based statistical inference in linear structural models with possibly weak instruments, *Econometrica* **73**(4): 1351–1365.

Dufour, J.-M. and Taamouti, M. (2007). Further results on projection-based inference in IV regressions with weak, collinear or missing Instruments, *Journal of Econometrics* **139**(1): 133–153.

Eichenbaum, M. and Fisher, J. D. M. (2007). Estimating the frequency of price re-optimization in Calvo-style models, *Journal of Monetary Economics* **54**(7): 2032–2047.

Gabriel, V., Levine, P., Spence, C. and Yang, B. (2008). On the (ir)relevance of direct supply-side effects on monetary policy, *Discussion Papers in Economics 04/08*, University of Surrey.

Galí, J. and Gertler, M. (1999). Inflation dynamics: A structural econometric analysis, *Journal of Monetary Economics* **44**(2): 195–222.

Galí, J., Gertler, M. and López-Salido, J. D. (2001). European inflation dynamics, *European Economic Review* **45**(7): 1237–1270.

Ghysels, E. and Hall, A. (1990). Testing nonnested Euler conditions with quadrature-based methods of approximation, *Journal of Econometrics* **46**(3): 273–308.

Hahn, J. and Hausman, J. (2002). A new specification test for the validity of instrumental variables, *Econometrica* **70**(1): 163–189.

Hall, A. R. (2005). *Generalized Method of Moments*, Oxford University Press, Oxford.

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

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators, *Econometrica* **50**(4): 1029–1054.

Hansen, L. P., Heaton, J., Lee, J. and Roussanov, N. (2007). Intertemporal substitution and risk aversion, *in* J. Heckman and E. Leamer (eds), *Handbook of Econometrics*, Vol. 6 of *Handbook of Econometrics*, Elsevier, chapter 61.

Hansen, L. P., Heaton, J. and Yaron, A. (1996). Finite sample properties of some alternative GMM estimators, *Journal of Business & Economic Statistics* **14**(3): 262–280.

Huelsewig, O., Henzel, S., Wollmershaeuser, T. and Mayer, E. (2009). Price puzzle revisited: Can the cost channel explain a rise in inflation after a monetary policy shock?, *Journal of Macroeconomics* **31**(2): 268–289.

Judge, G. G., Griffiths, W., Hill, R. C., Luetkepohl, H. and Lee, T.-C. (1985). *The Theory and Practice of Econometrics*, second edn, John Wiley and Sons.

Kapetanios, G. and Marcellino, M. (2010). Factor-GMM estimation with large sets of possibly weak instruments, *Computational Statistics & Data Analysis* **54**(11): 2655–2675.

Kleibergen, F. (2007). Generalizing weak instrument robust IV statistics toward multiple parameters, unrestricted covariance matrices and identification statistics, *Journal of Econometrics* **139**(1): 181–216.

Kleibergen, F. and Mavroeidis, S. (2009a). Inference on subsets of parameters in GMM without assuming identification, *Technical report*, Brown University.

Kleibergen, F. and Mavroeidis, S. (2009b). Weak instrument robust tests in GMM and the New Keynesian Phillips curve, *Journal of Business and Economic Statistics* **27**(3): 293–339.

Kocherlakota, N. R. (1990). On tests of representative consumer asset pricing models, *Journal of Monetary Economics* **26**(2): 285–304.

Kuester, K., Müller, G. and Stölting, S. (2009). Is the New Keynesian Phillips curve flat?, *Economics Letters* **103**(1): 39–41.

Ma, A. (2002). GMM estimation of the new Phillips curve, *Economics Letters* **76**(3): 411–417.

Martins, L. F. and Gabriel, V. J. (2009). New Keynesian Phillips curves and potential identification failures: A Generalized Empirical Likelihood analysis, *Journal of Macroeconomics* **31**(4): 561–571.

Mavroeidis, S. (2005). Identification issues in forward-looking models estimated by GMM with an application to the Phillips curve, *Journal of Money, Credit and Banking* **37**(3): 421–449.

Mavroeidis, S. (2007). Testing the New Keynesian Phillips curve without assuming identification, *Working paper*, Brown University.

Moreira, M. J. (2003). A conditional likelihood ratio test for structural models, *Econometrica* **71**(4): 1027–1048.

Nason, J. M. and Smith, G. W. (2008). Identifying the new Keynesian Phillips curve, *Journal of Applied Econometrics* **23**(5): 525–551.

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* **55**(4): 703–708.

Newey, W. K. and West, K. D. (1994). Automatic lag selection in covariance matrix estimation, *Review of Economic Studies* **61**(4): 631–653.

Rabanal, P. (2007). Does inflation increase after monetary policy tightening? Answers based on an estimated DSGE model, *Journal of Economic Dynamics & Control* **31**(3): 906–937.

Ravenna, F. and Walsh, C. E. (2006). Optimal monetary policy with the cost channel, *Journal of Monetary Economics* **53**(2): 199–216.

Scheufele, R. (2010). Evaluating the German (New Keynesian) Phillips curve, *The North American Journal of Economics and Finance* **21**(2): 145–164.

Singleton, K. J. (1985). Testing specifications of economic agents' intertemporal optimum problems in the presence of alternative models, *Journal of Econometrics* **30**(1-2): 391–413.

Smets, F. and Wouters, R. (2003). An estimated dynamic stochastic general equilibrium model of the Euro area, *Journal of the European Economic Association* **1**(5): 1123–1175.

Smith, R. J. (1992). Non-nested tests for competing models estimated by Generalized Methods of Moments, *Econometrica* **60**(4): 973–980.

Stock, J. H. and Watson, M. W. (1999). Forecasting Inflation, *Journal of Monetary Economics* **44**(2): 293–335.

Stock, J. H. and Watson, M. W. (2002). Macroeconomic forecasting using diffusion indexes, *Journal of Business and Economic Statistics* **20**(2): 147–162.

Stock, J. H. and Wright, J. H. (2000). GMM with weak identification, *Econometrica* **68**(5): 1055–1096.

Stock, J. H., Wright, J. H. and Yogo, M. (2002). A survey of weak instruments and weak identification in generalized method of moments, *Journal of Business & Economic Statistics* **20**(4): 518–29.

Stock, J. and Yogo, M. (2005). Testing for weak instruments in linear IV regression, *in* D. W. K. Andrews and J. H. Stock (eds), *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, Cambridge: Cambridge University Press.

Tauchen, G. (1986). Statistical properties of generalized method-of-moments estimators of structural parameters obtained from financial market data, *Journal of Business and Economic Statistics* **4**(4): 397–416.

Tillmann, P. (2008). Do interest rates drive inflation dynamics? An analysis of the cost channel of monetary transmission, *Journal of Economic Dynamics and Control* **32**(9): 2723–2744.

Walsh, C. E. (2003). *Monetary theory and policy*, 2nd edn, The MIT Press, Cambridge, Massachusetts.

West, K. D. (1997). Another heteroskedasticity- and autocorrelation-consistent covariance matrix estimator, *Journal of Econometrics* **76**(1-2): 171–191.

Woodford, M. (2003). *Interest and Prices: Foundations of a Theory of Monetary Policy*, Princeton University Press, Princeton.

Wright, J. H. (2009). Comments on "Weak instrument robust tests in GMM and the New Keynesian Phillips curve" by F. Kleibergen and S. Mavroeidis, *Journal of Business and Economic Statistics* **27**(3): 323–326.

Wright, J. H. (2010). Testing the adequacy of conventional asymptotics in GMM, *Econometrics Journal* **13**(2): 205–217.

A Appendix

A.1 Determining the appropriate HAC estimator

Determining the estimated long-run covariance matrix $\hat{S}_T$ is also crucial for inference, particularly when heteroskedasticity and/or autocorrelation are present. The class of HAC estimators has therefore been proposed. These estimators may be classified into two broad categories: non-parametric kernel-based procedures and parametric procedures.

The most frequently applied method is Newey and West (1987), a kernel-based procedure. In this case, $\hat{S}_T$ would be

$$\hat{S}_T^{NW} = \hat{\Gamma}_{0,T} + \sum_{v=1}^{q} (1 - [v/(q+1)]) \left(\hat{\Gamma}_{v,T} + \hat{\Gamma}'_{v,T} \right)$$

where

$$\hat{\Gamma}'_{v,T} = (1/T) \sum_{t=v+1}^{T} \left[f_t(\hat{\vartheta}) \right] \left[f_t(\hat{\vartheta}) \right]' .$$

With $f_t(\hat{\vartheta})$ denoting the k orthogonality conditions for this model which depend on the parameters and the data. Note that $\hat{\vartheta}$ and $\hat{S}_T$ are estimated jointly owing to the application of the CUE.

A critical aspect of this approach is the determination of the optimal bandwidth q. Andrews (1991) and Newey and West (1994) have therefore proposed selection rules for determining the optimal bandwidth. In addition Andrews and Monahan (1992) and Newey and West (1994) promoted the construction of pre-whitened and re-colored HAC estimators to provide a better finite sample performance. However, neither of these more advanced procedures works without problems in our application, and endogenous selection leads to an implausibly high value of q for each of the selected methods (regardless of whether pre-whitening is employed).

As an additional method of HAC estimation parametric approaches exists (see Den Haan and Levin, 1997, for an overview). Given the model structure employed here with our instrument set, we can assume that the residuals are generated by a specific parametric model. Since we use only instruments dated $t-1$ or earlier, we can use the procedure of West (1997), using the MA(1) structure. More specifically, we can define

$$\hat{u}_t = h_t(\hat{\vartheta}),$$

where $\hat{u}_t$ is given by the model structure (see eq. 9). Then let $\hat{\theta}_1$ be a consistent estimator of θ_1 based on $\hat{u}_t = \hat{\epsilon}_t + \hat{\theta}_1 \hat{\epsilon}_{t-1}$. We determine $\hat{\epsilon}$'s and $\hat{\theta}_1$ by applying a weighted nonlinear least squares procedure (see e.g., Brockwell and Davis, 1990, chapter 8.7). Then we define the vector $\hat{d}_{t+1}$ as

$$\hat{d}_{t+1} = \left(\mathbf{z}_{t-1} + \mathbf{z}_t \hat{\theta}_1\right) \hat{\epsilon}_t$$

and can estimate S_T as

$$\hat{S}_T^{MA} = (T-1)^{-1} \sum_{t=1}^{T-1} \hat{d}_{t+1} \hat{d}_{t+1}^{\prime}.$$

$\hat{S}_T^{MA}$ is positive semidefinite per construction.

West (1997) shows that this method works well in situations where cross-products of instruments and disturbances are sharply negatively correlated. This is also the situation where "truncated" (non-parametric) estimators are likely to fail. This seems to be the case for us, because θ is estimated to be around -0.6. We also calculated the residuals of the estimated model and ran some diagnostic tests for autocorrelation, so we can confirm the MA(1) process. The implication is that we cannot confirm the existence of a cost push shock as stressed by many DSGE models estimated under full information, which are assumed to be positively correlated (mostly following an AR(1) process). To allow for more flexible specifications of the error term, we also consider Newey and West's (1987) standard procedure, whereby the bandwidth is set up by rule of thumb $q = int\left\{T^{1/3}\right\}$.

A.2 MCMC

The MCMC simulations follow a version of the standard Metropolis-Hastings algorithm (see Chernozhukov and Hong, 2003; Hansen et al., 2007). Let the parameter combination corresponding to the ith draw be $b^{(i)} = \left[\beta^{(i)}, \theta^{(i)}, \xi^{(i)}\right]$, given the calibrated parameters α, ϵ and ϕ. Then

1. Take $b^{(0)}$ as the CUE point estimator $b^{(0)} = b^{(CUE)}$.
2. Draw ς from the conditional distribution $q(\varsigma|b^{(i)})$.
3. With probability $\inf\left(\frac{\exp(-\mathbf{S}(b^{(i+1)}))q(b^{(i)}|\varsigma)}{\exp(-\mathbf{S}(b^{(i)}))q(\varsigma|b^{(i)})}, 1\right)$ update $b^{(i+1)} = \varsigma$; otherwise keep $b^{(i+1)} = b^{(i)}$.

We take a Gaussian transition density, which results in a Markov chain, which is a random walk. We also constrain the parameter space to match the economically plausible range between 0 and 1 (which guarantees a compact set). Further, we allow only parameter combinations to include that pass the χ^2 test; i.e. $T \times \mathbf{S}(b) <$

$c_k(\alpha)$ (see eq. 13). Let ϕ be the trivariate normal density centered around zero with cdf Φ. Then

$$q(x|y) = \frac{\phi(x-y)}{\Pr(x \in A)}, \quad \text{where } x = y + z,\ z \sim \Phi,$$

where A is the event that b falls outside its boundaries or the χ^2 test rejects. In the simulation, the truncation is accomplished by discarding those values that do not fulfill these requirements. A choice has to be made concerning the dispersion of ϕ (the different variances). We take different values for each parameter in order to obtain an acception rate of the algorithm that matches a value slightly above 25%. The reported results are based on simulations with 1,000,000 accepted draws.

A.3 Test of adequacy of convential asymptotics in GMM

Wright (2010) proposed a test for the null hypothesis that a GMM model is sufficiently well identified for conventional asymptotics to be reliable. This test can be applied for various Phillips curve specifications in order to decide whether it is necessary to rely on more robust procedures, which could also be associated with a power loss.

Setting up the test statistic as proposed by Wright (2010), we define W_1 as the maximum distance between any two values of ϑ in the **S**-set for ϑ. If the **S**-set is unbounded, W_1 will be infinite. When ϑ is a vector, ϑ_1 and ϑ_2 are the lowest and largest values, obtained respectively from the projection method and W_1 will be the value that maximizes these distances.

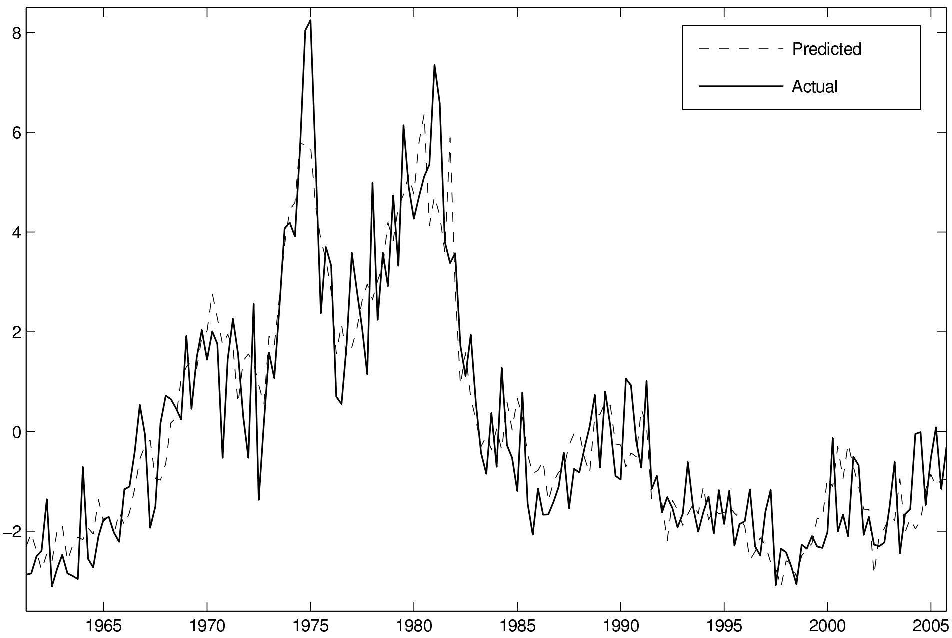
Likewise, W_2 can be defined as the maximum between any two points in the usual 2GMM Wald confidence set for ϑ, which is given by $T(\hat{\vartheta}_{TS} - \hat{\vartheta}_i)' \hat{J} (\hat{\vartheta}_{TS} - \hat{\vartheta}_i) \leq c_n(\alpha)$ for $i = 1, 2$, where $c_n(\alpha)$ is the $100(1-\alpha)\%$ percentile of the χ^2 distribution. Numerically, W_2 can be computed as $W_2 = \frac{2}{\sqrt{T}} \sqrt{\frac{c_n(\alpha)}{\hat{\lambda}}}$, where $\hat{\lambda}$ is the smallest eigenvalue of $\hat{J}$. The test then simply compares the two volumes of the confidence set equal to

$$L = \frac{W_1}{W_2}.$$

The author provides critical values for this test statistic (see Wright, 2010, Table 1).

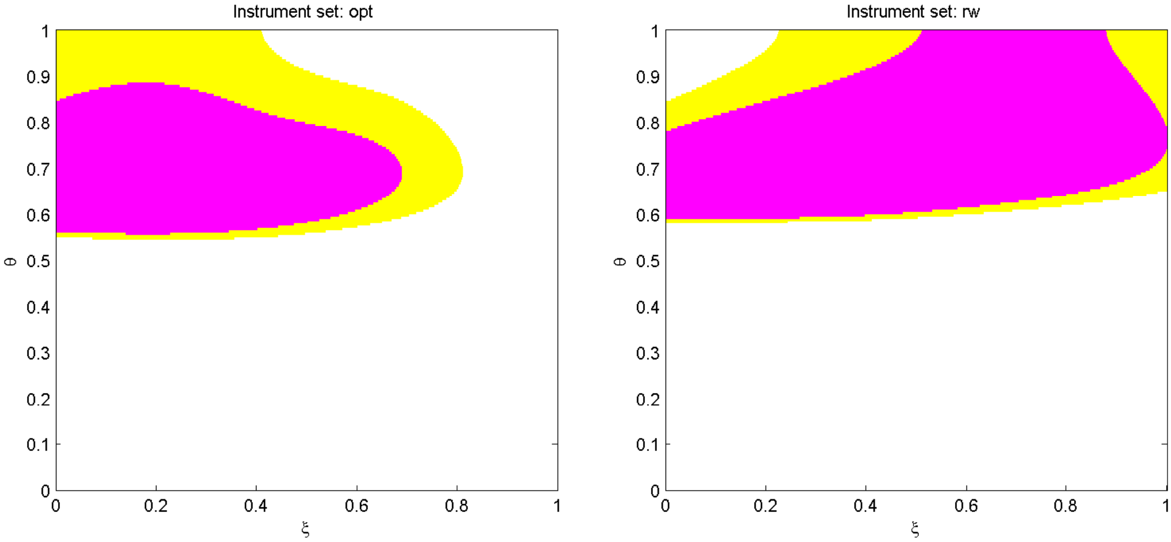
A.4 Figures

Figure 4: Fit of the pure forward-looking model with the cost channel
(Actual vs. predicted Inflation)



Notes: Actual and predicted annualized demeaned inflation rates based on the GDP deflator.

Figure 5: **S**-sets for the structural parameters



Notes: Joint confidence sets on the structural parameters (θ,ξ) in the NKPC model with a cost channel given the calibrated values $\beta = 0.99$ and $\phi^l = 1$. Different instrument sets used. The shaded areas contain joint 95% (yellow) and 90% (magenta) confidence sets.